MIKON 94

X INTERNATIONAL MICROWAVE CONFERENCE
POLAND, KSIĄŻ, MAY 30 - JUNE 2, 1994

AD-A284 353

TELECOMMUNICATIONS RESEARCH INSTITUTE

10TH INTERNATIONAL MICROWAVE CONFERENCE

MIKON - 94

INVITED PAPERS

94-29129

KSIĄŻ CASTLE – POLAND

May 30 – June 2, 1994

VOLUME 3

CONTENTS

The symbol before title of paper is identicaly as in conference program and shows session, in which this paper appear.

The paper with missing numbers were not available before deadline.

NEW TRENDS AND IDEAS IN THE FIELDS
OF
MICROWAVE TECHNOLOGY

Tatsuo Itoh*

ABSTRACT

This paper presents a subjective view toward new directions and some examples for changing microwave research. The paper is written primarily from the point of view of how the electromagnetic research must change and what kind of impact such a change can give rise to stimulation for the device and circuit research for microwave technology. It is emphasized that interdisciplinary treatment of electromagnetic research is vital for the future of microwave technology.

1. INTRODUCTION

Recently, the microwave and "radio" engineering has drawn a renewed interest due to a significant interest in personal and mobile communications and intelligent automobile And others. At the same time, due to the end of the cold war era, the traditional use of microwave is changing to dual use from military use. Expanded role of microwave technologies to civilian and commercial use require some changes in design philosophy and fabrication methods of microwave devices, circuits, components and subsystems. Accurate first-pass design is now more important. Packaging must be integral part of the circuit and component design to reduce the assembly cost. Working environment of these commercial-use components is as harsh as or even more harsh than the military components.

It is of no doubt that the research on new devices and new fabrication method will be enhanced. As one climbs the "food chain" of the microwave technology, the new system design responsive to the need for the commercial application will emerge. Traditionally, the circuit design filled the link between these two extremes of the food chain. Here, the role of electromagnetics has been well understood.

As the key element of the microwave technology, this paper addresses the

---

* Electrical Engineering Department, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90024-1594, U. S. A.

changing need for the electromagnetics. It is well appreciated that the electromagnetics have been a fundamental subject of study but is essential to microwave circuit and component design. Nevertheless, the research on electromagnetics, especially the one exercised in the academic community has not been very responsive, or not sufficiently so, to the practical problems. It is necessary to expand the role of electromagnetic research by making it more interdisciplinary. By this, it is meant that the traditional electromagneic research must be evolved into a "system" science by including the wave interaction aspects of the circuits with active devices, packaging effect and the visualization oriented analysis and design. By means of examples, some of the effort along this line will be explained in this paper. Appropriate exploitation of such an approach is believed to reach a new plateau for the modem electromagnetics from which new and enhanced capability of microwave technology will emerge.

2. INTERACTION WITH ACTIVE DEVICES

The interaction of the electromagnetic wave with active devices needs to be considered carefully and must be understood well in the new microwave circuits which are increasingly complex. In many cases, the passive and active part of the circuit cannot be distinguished clearly. In such cases, it is important to characterize the entire circuit with a single design process. For such a purpose, the enhanced electromagnetics should play an important role. The wave-device interaction must be taken into account. In addition to the comprehensive design approach, a new type of circuit/component may be created. An example is the active antenna. The antenna acts not only as a radiating element but also provides a circuit function such as diplexer or resonator. This latter function becomes an integral part of the active circuit such as the oscillator[1,2]. Fig.1 shows an example of active antenna. The patch antenna is not only a radiating element but also a resonant feedback element for the active device to form an oscillator. The structure is a two-element array. In this arrangement, the radiation pattern can be switched from the difference pattern to the sum pattern. Because the two oscillators are strongly coupled by a connecting microstrip line, they support two oscillation modes under a small signal condition. However, due to the large signal nonlinear behavior of the circuit, only one of them becomes a stable oscillation mode under steady state condition. As to which mode becomes stable has been a subject of study in [3].

It is well known that the antenna is one of the bottleneck issues for personal communication. The concept of an active antenna can add some "intelligent" function right at the antenna. The design and analysis of this type of circuits must be enhanced.
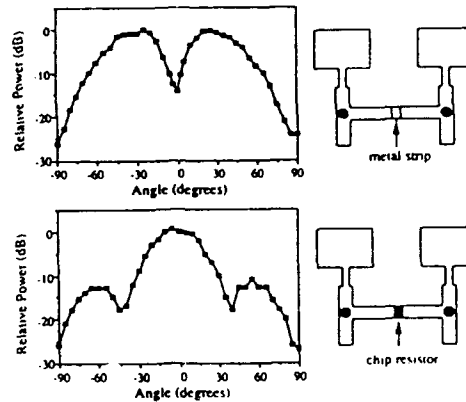
Fig.1   Two-element active antennas

3. UNIFIED ELECTROMAGNETIC CHARACTERIZATIONS OF MMIC

State of the art microwave and millimeter wave circuits are becoming more complex. They perform multiple tasks, are electrically smaller and are more dense. In some cases, the circuits are multimoded. In many cases, the current computer aided design tools are unable to correctly predict the circuit behaviors. In most of the microwave integrated circuit design, the passive part and the active part are treated differently. In the last several years, a significant advance has been reported in the area of computer aided design (CAD). Advances in the passive part now allows EM simulators which incorporate electromagnetic behaviors of the passive circuit instead of classical quasi-static analysis. Electromagnetic coupling between nearby elements can be included correctly and increasingly the environmental effects such as packaging and multi-chip interconnect can be taken into account. The interactions of the passive part with the active devices, however, have been carried out typically by a circuit simulator such as the harmonic balance method. Therefore, electromagnetic interactions with the passive and active elements of the circuits may not be correctly included. In the next generation of the CAD, it is necessary that all of the effects including the packaging effects to the active devices and the electromagnetic coupling between the active devices are taken into account correctly.

Most of the frequency domain analysis methods for the passive structure cannot easily handle active devices in the dynamic mode. Certain approximate steady state

analysis can be made, however. On the other hand, the time domain algorithm such as the TLM (Transmission Line Matrix) method and the FDTD (Finite-Difference Time-Domain) method can include the active device in its dynamic (transient) as well as steady state operations. This is due to the fact that the time domain method is an initial value problem and hence can include the build up of the nonlinear device operation in the time evolving history until the steady state is reached. Hence, by the use of the time domain algorithm, it is possible to simulate the entire microwave circuit including the nonlinear active device and the packaging effect. Here, the latter is interpreted in a broad sense.

A recent effort by the author's group has produced a remarkably interesting results for a number of microwave circuits including the active devices. An example is shown in Fig.2. By means of the modified FDTD, the correct oscillation mode is established for

Example: Simulation of an active antenna.
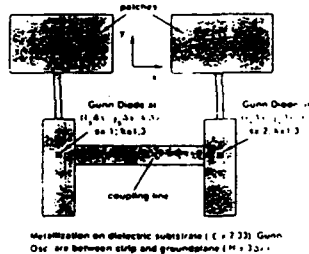
circuit
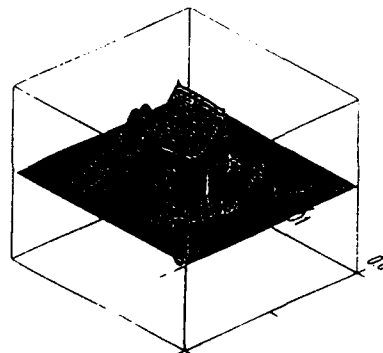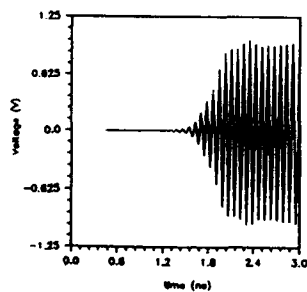
Anti- Phase Mode
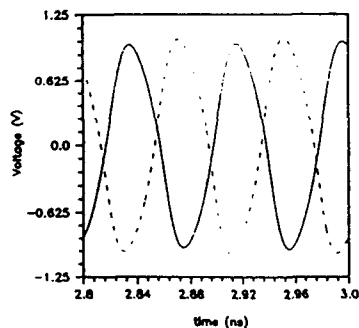
Simulation predicts correct steady state oscillation mode.

Fig.2   Active antenna characterized by enhanced FDTD

the coupled active antenna shown in Fig.1. The visualization aspect of the FDTD is particularly a welcome addition for simulation, because the defective design can be exhibited and the design concept can be visually verified.

In contrast to many frequency domain techniques such as the integral equation technique which has a numerical efficiency at the expense of the high analytical labor, the time domain methods typically require a small amount of the analytical processing. Instead, a large memory and a long computation time are needed [4,5]. Also, the inputting the structural data and the extraction of the needed design parameters are where effort must be spent. The latter has an advantage that it is possible to obtain more information such as the dynamic field behavior than the simple circuit parameters.

As the computational power is increasing, the computational burden is being reduced. Nevertheless, a significant effort has been expended to make the time domain method algorithmically and computationally more efficient. Diakoptics and the system identification are two of such examples. In the first, a large linear passive portion is replaced with its characteristics at the "terminal" plane. By means of the Diakoptics or numerical Green's function indicating the impulse response at that terminal plane, the computational volume can be reduced [6]. This concept is useful in design of a large circuit in which a small portion is tuned but the large portion unchanged. In addition, the concept can make the calculation more efficient in the circuit which contain an active device. The terminal plane of the active circuit can be made to interact with the device. On the other hand, the system identification technique is a prediction method based on the technique developed for the linear system theory [7].

Due to its algorithm, the time domain methods such as FDTD are suited for parallel processing. Although a parallel processing machine can be used, the parallel algorithm can be implemented into a parallel virtual machine (PVM) made up of a cluster of workstations.

4. PACKAGING, INTERCONNECT AND MULTIMODING EFFECTS

Recently, importance of the effect of packaging has been recognized by the circuit designer. In many cases, the packaging effect implies the deviation of the designed circuit performance when it is placed in a package. However the meaning of the packaging should be interpreted in a wider sense. For instance, a multilayered circuit as shown in Fig.3 can be considered as the case in which the packaging is an integral part of the circuit design [9]. In a highly complicated circuit configurations, possible "multi mode" situations may exist. Such modes may include undesirable ones such as the ones
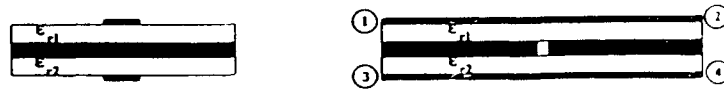
Fig.3 Multilayered interconnect/packaging structure

that are leaky. One of the highly publicized phenomena is the lateral leakage in a conductor backed slot line of coplanar waveguide[8]. This leakage is due to the parallel plate mode and is now well understood. Although this leakage is detrimental in many cases, such a leakage can be advantageously used for constructing a new type of passive device [10]. More recently, a new method for controlling such a leakage by means of a multilayered structure has been invented [10].

5. CONCLUSIONS

A subjective view has been presented for the new trends and ideas for microwave technology from the point of view of enhanced electromagnetics. This concept is believed useful for advancing state of the art of microwave circuits and components design. Several examples have been described.

ACKNOWLEDGMENT

REFERENCES

[1] J. Birkeland and T. Itoh, "Planar FET oscillators using periodic microstrip patch antennas," IEEE Trans. Microwave Theory and Techniques, Vol. 37, No. 8, pp. 1232-1236, Aug. 1989.

[2] J. Lin, T. Itoh and S. Nogi, "Mode switch in a two-element active array," 1993 IEEE AP-S Symposium Digest, pp.664-667, June 28-July 2, 1993, Ann Arbor, MI.

[3] S. Nogi, J. Lin and T. Itoh, "Mode analysis and stabilization of a spatial power combining array with strongly coupled oscillators," IEEE Trans. Microwave Theory and Techniques, Vol.41, No. 10, pp.1827-1837, October 1993.

[4] K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," IEEE Trans. Antennas & Prop., vol. AP-14, No.,302-307, May 1966.

[5] W. J. R. Hoefer, "The transmission line matrix method - Theory and applications," IEEE Trans. Microwave Theory and Tech., Vol.MTT-33, No.10, pp.882-893, Oct. 1985.

[6] T. W. Huang, B. Houshmand and T. Itoh, "The FDTD Diakoptics method," IEEE MTT-S International Microwave Symp. Dig., pp.1435-1438, June 1993, Atlanta, GA.

[7] B. Houshmand and T. Itoh, "Microwave structure characterization by a combination of FDTD and system identification methods," IEEE Microwave and Guided Wave Letters, Vol.3, No.8, pp.262-264, August 1993.

[8] H. Shigesawa, M. Tsuji and A. A. Oliner, "Conductor-backed slot line and coplanar waveguide: dangers and full-wave analysis," IEEE MTT-S Int. Microwave Symp. Dig., pp.199-202, May 1988, New York, NY.

[9] A. M. Tran and T. Itoh, "Open-ended microstrip lines coupled through an arbitrary aperture in a thick common ground plane," 23rd European Microwave Conference, p.435, September 6-9, 1993, Madrid, Spain.

[10] Y. Liu, T. Hirota and T. Itoh, "Coupling phenomena in conductor-backed slotline structures," IEEE MTT-S Int. Microwave Symp. Dig., pp.1203-1206, Atlanta, GA.

[11] Y. Liu and T. Itoh, "Leakage phenomena in multilayered conductor-backed coplanar waveguide," IEEE Microwave and Guided Wave Letters, Vol.3, No. 11, pp.426-427, November 1993.

DEVELOPMENT OF A DUAL-REFLECTOR FEED FOR THE ARECIBO RADIO TELESCOPE, AN OVERVIEW

Per-Simon Kildal*

Abstract

The spherical reflector antenna of the Arecibo Observatory will be equipped with a dual-reflector feed, referred to as the Gregorian. The Gregorian has larger band-width, lower losses and provides better aperture illumination than the old line feeds. This presentation reviews the methods and algorithms that have been developed to synthesize the shapes of the two Gregorian feed reflectors, and to analyze them. The analysis is done both by forward ray tracing using geometrical optics with included edge diffraction corrections, and by physical optics (PO) integration. The horn feeds and the Gregorian geometry are optimized to improve the field-of-view and reduce the weight. A "mini-Gregorian" has already been constructed, built and tested to verify the dual-reflector feed concept.

I.   INTRODUCTION

The main instrument of the Arecibo Observatory is a spherical reflector antenna with an aperture diameter of 305 m. This antenna is used in the frequency range from 25 MHz to 2.5 GHz for radio astronomy, for radar investigations of the solar system and for ionosphere research using incoherent scatter radar. The spherical reflector is presently fed by line feeds which correct for spherical aberration. All of the line feeds suffer from three major deficiencies: i) Ohmic losses which increase the noise temperature and hence, reduce the sensivity of the radio-telescope. ii) Limited bandwidth which requires new line feeds when new frequency bands are needed, a laborious and expensive process. iii) Under-illumination of the central region of the reflector causing high sidelobe in the radiation pattern.

To overcome these major drawbacks, Cornell University, which operates the Observatory, is about to replace the line feeds by a dual-reflector feed, consisting of two subreflectors and several feed horns. These subreflectors have concave shapes and resemble the ellipsoidal subreflector of the classical Gregorian telescope. The dual-reflector feed is, therefore, often referred to as the "Gregorian" feed.

* Department of Microwave Technology
Chalmers University of Technology
S-412 96 Gothenburg, Sweden

The purpose of the present paper is to summarize the work for the Arecibo upgrading which has gone into the design of the dual-reflector feed. A more complete summary is already given in [1], and [2], where [2] is most complete and updated. Details of this work have already been presented at conferences and documented in journal articles; i.e. methods to synthesize the shapes of the two reflectors [3]-[5], methods to analyse the geometrical optics (GO) performance [6]-[7] and the edge diffraction losses [8]-[9], verification of the theoretical shapes by ray tracing and physical optics analysis [10], and experimental verification of the concepts by the construction and testing of the "mini-Gregorian" on the telescope, see Figure 1 and [1]-[2]. The work has also involved optimization of the reflector shapes with respect to mechanical constraints such as overall size, weight, self-blockage and horn sizes. Multiple beams, field of view, and tolerances were also studied [11].

II.   THE GREGORIAN DUAL-REFLECTOR FEED

The spherical reflector of the Arecibo antenna is fixed to the ground, facing toward zenith. It has a diameter of 305 m (1000 ft) and a radius of curvature of 265 m (870 ft). The central rays cross at the paraxial focus located half way between the center of the sphere and the surface. All the other rays cross the symmetry axis below this point along a focal line. The line feeds are located along the focal line to combine the rays with correct amplitude and phase.

The Gregorian dual-reflector feed combines the incoming rays by focussing them to a focal point, in which a feed horn is located (Figure 2). The two Gregorian feed reflectors are referred to as the secondary and the tertiary, and the spherical reflector is the primary. The feeds are mounted under a feed arm which rotates from a large triangular structure suspended by cables. Movement of the Gregorian along the feed arm provides elevation steering, and rotation of the feed arm provides azimuth steering, in the same way as for the line feeds. THe illuminated aperture over the primary is not located symmetrically around the feed axis. In fact, there is an offset of 15 m (50 ft) between this axis and the center of the aperture, which makes the zenith angle where spillover occurs larger than with a symmetrical aperture. The illuminated aperture is increased without added spillover by making the aperture elliptical in the direction transverse to the elevation steering. The aperture diameter in the symmetry plane is 213 m (700 ft), whereas the transverse diameter is increased to 237 m (770 ft). The radiation pattern will have a slightly elliptical shape but this is acceptable for almost all radio astronomy applications.

III. SYNTHESIS

The Arecibo dual-reflector feed is designed by using modern geometrical optics (GO) to provide an efficient illumination of the aperture of the sphere, having constant phase and low cross-polarization. The laws of modern GO is not only the classical reflection and power conservation laws most commonly used in reflector synthesis, but also: i) the "GO continuation or transmission formula" for the amplitude, phase and polarization of the E-field; ii) the "field reflection law"; and iii) the "lens formula for reflection" from which the local principal directions and principal curvatures of the reflector can be calculated from the principal directions and principal curvatures of the wavefronts of the incident and reflected fields, see [5], [7] and [3]. By using all these laws as described in [3] and [4], it has been possible to reduce the synthesis problem to solving a number of linear non-differential equations in series. This makes it easier to control the result, and an exact and accurate solution is obtained without approximations.

IV. ANALYSIS BY FORWARD RAY TRACING

The dual-reflector feed has been analyzed extensively by GO ray tracing (high frequencies) and by physical optics integration (low frequencies). An algorithm and a corresponding computer code for forward GO ray tracing has been developed [7]. Forward ray tracing (FRT) means that a bundle or rays are launched from the horn feed, and these rays are followed via reflection from each reflector till they end in the aperture plane. The code can trace rays via an arbitrary number or numerically specified reflectors. The FRT code has been extensively used in the tolerance studies, and in designing feed clusters at high frequencies [11].

V. EDGE DIFFRACTION CORRECTION

Using forward ray tracing, edge diffraction can be included very efficiently by using the transition region theory (TRT) [8]-[9] without tracing additional rays. The TRT approach is based on the fact that the main effect of edge diffraction in GO designs is to "smear-out" the sharp GO boundaries. In other words, edge diffraction gives the total field at the GO boundary a finite slope caused by the so-called transition region. By the TRT, we can calculate the lateral extent $\Delta\rho$ of this transition region from the GO wavefront curvatures at the GO boundaries. From this $\Delta\rho$, we can construct a standard field behaviour around the boundary, which is valid inside the transition region and near [8]. By tracing $\Delta\rho$ to the aperture, the aperture integral as well as the diffraction spillover can be calculated by

evaluating a line integral around the aperture [9]. Thereby, the efficiency reduction due to edge diffraction is found.

VI.   PHYSICAL OPTICS INTEGRATION

Multi-reflector antennas can also be analyzed by straight-forward physical optics (PO) integration. However, this requires an enormous amount of computer time t if the wavelength λ is small, as $t \propto 1/\lambda^4$. The PO analysis of the Arecibo feed has been done by an efficient PO code. PO can be used with satisfactory accuracy to analyze the Gregorian up to about 1 GHz. Above 1 GHz, the computer time becomes too large.

The calculated PO field in the aperture is in good agreement with the GO field at 1 GHz. There are two main discrepancies; the PO field has a finite slope at the GO boundary and a dip in the center of the aperture. The slope is due to edge diffraction and agrees with that obtained by the TRT, see [10, Sec 4]. The dip is more pronounced at low frequencies. It is centered around the z axis at $\rho=50$ feet and not around the center of the illuminated aperture. The dip is actually a diffraction effect associated with the paraxial focus (i.e. the cusp of the caustic surfaces meeting there), and we therefore refer to it as cusp diffraction. The effect is studied in detail in [10, Sec. 5]. The dip is reduced by lifting the dual-reflector feed, in such a way that the spacing between its surface and the paraxial focus is increased.

VII.   APERTURE BLOCKAGE AND EFFICIENCY

The aperture efficiency is shown in Figure 3. It is calculated for the actual case with the 305 m diameter primary, and the center of the aperture is assumed to be located in the center of the primary. The efficiency is referred to a uniformly illuminated elliptical aperture with 213 m x 237 m diameters.

The subefficiencies shown are a result of a factorization based on different diffraction phenomena:

i) The GO efficiency is the contribution due to spillover outside the tertiary reflector (-0.11 dB) and the specified illumination taper of the aperture obtained at high frequencies (-0.12 dB), when the tertiary is correctly illuminated by the feed horn.

ii) The subefficiency caused by the shadowing of the feed platform is estimated as explained below. The platform is assumed to be an 80% transparent circular disk, centered in the

aperture, with a diameter of 91 m. This gives a geometrical blockage of 3.3% for the Gregorian having an elliptical aperture with diameters of 213 m and 237 m (for the mini-Gregorian it will be 16%). The corresponding subefficiency becomes, for our nearly-uniformly illuminated aperture, -0.3 dB. There will be a corresponding increase in noise temperature due to scattering towards the ground of 2 or 3 Kelvins. The platform will also increase the sidelobes.

iii) The loss due to edge diffraction is found in the FRT-code from the formulas in [8] and [9]. This loss varies with the square root of the wavelength.

iv) The loss due to cusp diffraction, i.e. due to the illumination dip in the aperture at low frequencies, is found as the efficiency resulting from the PO calculation minus contributions i) and iii) above.

v) The loss caused by the surface tolerances are readily found from Ruze's formula, using 2.3 mm rms of the spherical primary.

The degradations due to diffraction shadowing by the feed housing, the two reflectors and the opening of the enclosure (that will surround the Gregorian) are also indicated in Figure 3. These effects will give efficiency reductions below 430 MHz, see [12] for more details. All efficiencies are based on a correct illumination of the tertiary with an ideal $\cos^n(\theta/2)$ feed pattern, providing 15 dB edge taper. In practice, this will not be ideal and will vary slightly with frequency away from the ideal pattern. This may introduce gain reductions of about 0.3 dB from the predictions.

VIII. CONCLUSIONS

Construction of the Gregorian dual-reflector feed for the Arecibo telescope begun in early 1993 and should be completed in 1995. The predictions show that it will be a valuable research tool, with continuous frequency coverage from below 430 MHz to above 8 GHz, increased aperture efficiency, lower system noise temperature, and multi-beam capability. The entire frequency range of 300 MHz to 8 GHz can be covered using only 5 to 10 interchangeable corrugated feed horns. The ohmic losses will be minimal because the reflector surfaces correcting the phase aberrations are essentially lossless, unlike the waveguide in the line feeds.

When completed the capability of the Arecibo Observatory will have been greatly enhanced. The upper operating frequency of the Observatory will have been moved from about 2.5 GHz

to 8 GHz, the instantaneous bandwidth will have been increased from less than 50 MHz to being controlled only by the horn design, the sensitivity of radio astronomy observations will have been improved by a factor of 2 to 4 and the sensitivity for radar astronomy observations will have been increased by factors of 10 to 40, depending on what object is being observed.

The techniques that was developed to synthesize and analyze the Gregorian Arecibo feed are general and can also be applied in other antenna design, see e.g. [13].

REFERENCES

[1] P-S. Kildal, L. Baker, and T. Hagfors, "Development of a dual-reflector feed for the Arecibo telescope: An overview", IEEE Trans. Antennas Propag., AP-33, pp. 12-17, 1991.

[2] P-S. Kildal, L. Baker and T. Hagfors, "The Arecibo Upgrading: Electrical design and expected performance of the dual-reflector feed system", to appear in a special issue of Proceedings of the IEEE, May 1994.

[3] P-S. Kildal, "Synthesis of multi-reflector antennas by kinematic and dynamic ray tracing", IEEE Trans. Antennas Propagat., AP-38, pp. 1587-1599, 1990.

[4] P-S. Kildal, "A new approach to the synthesis of reflector antennas", Radio Science, 26(2), pp. 619-623, 1991.

[5] P-S. Kildal, "Principles and conclusions of synthesis of reflector antennas", lecture given at 17th QMW Antenna Symposium, 11-12 April 1991, Lecture Notes, Queen Mary & Westfield College, London.

[6] P-S. Kildal, "Laws of geometrical optics mapping in multi-reflector antennas with application to elliptical apertures", IEE Proceedings, Part H, MOA, 6, pp. 45-453. 1989.

[7] P-S. Kildal, "Analysis of numerically specified multi-reflector antennas by kinematic and dynamic ray tracing", IEEE Trans. Antennas Propag., AP-38, pp. 1600-1606, 1990.

[8] P-S. Kildal and J. Stamnes, "Asymptotic transition region theory for edge diffraction. Part 1: Tracing transition regions via reflectors", IEEE Trans. Antennas Propagat., AP-38, pp. 1350-1358, 1990.

[9] P-S. Kildal, "An asymptotic transition region theory for edge diffraction. Part 2: calculation of diffraction losses in multi-reflector antennas", IEEE Trans. Antennas Propag., AP-38, pp. 1359-1365, 1990.

[10] P-S. Kildal and S.A. Skyttemyr, "Diffraction analysis of a proposed dual-reflector feed for the spherical reflector antenna of the Arecibo Observatory", Radio Science, Vol. 24, pp. 601-617, Sept-Oct 1989.

[11] P-S. Kildal, M. Johansson, T. Hagfors and R. Giovanelli, "Analysis of a cluster feed for the Arecibo tri-reflector system using forward ray tracing and aperture integration", IEEE Transactions on Antennas and Propagation, Vol. 41, No. 8, pp. 1019-1025, Aug 1993.

[12] N. M. Johansson, P-S. Kildal, L. Baker, "Study of diffraction blockage effects in the Arecibo dual-reflector feed using physical optics", to be presented at IEEE AP-S Symposium, Seattle, June 1994.

[13] P-S. Kildal and J. Pezzani, "A dual-reflector feed for the radio telescope in Nancay synthesised by geometrical optics", to be presented at IEEE AP-S Symposium, Seattle, June 1994.
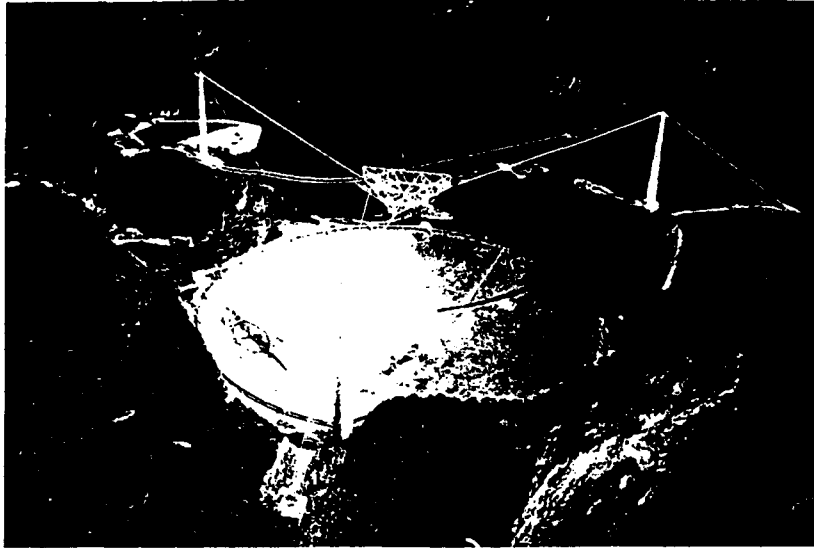
Figure 1a   The spherical reflector antenna of the Arecibo Observatory.
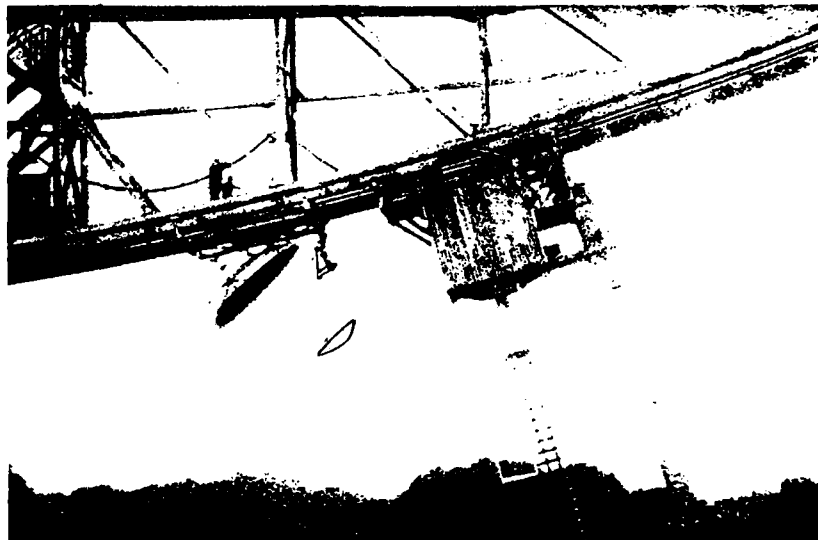
Figure 1b   Mini-version of the Gregorian feed under test in Arecibo, hanging beside two
of the line feeds. The full-size Gregorian will be four times as large, linearly.
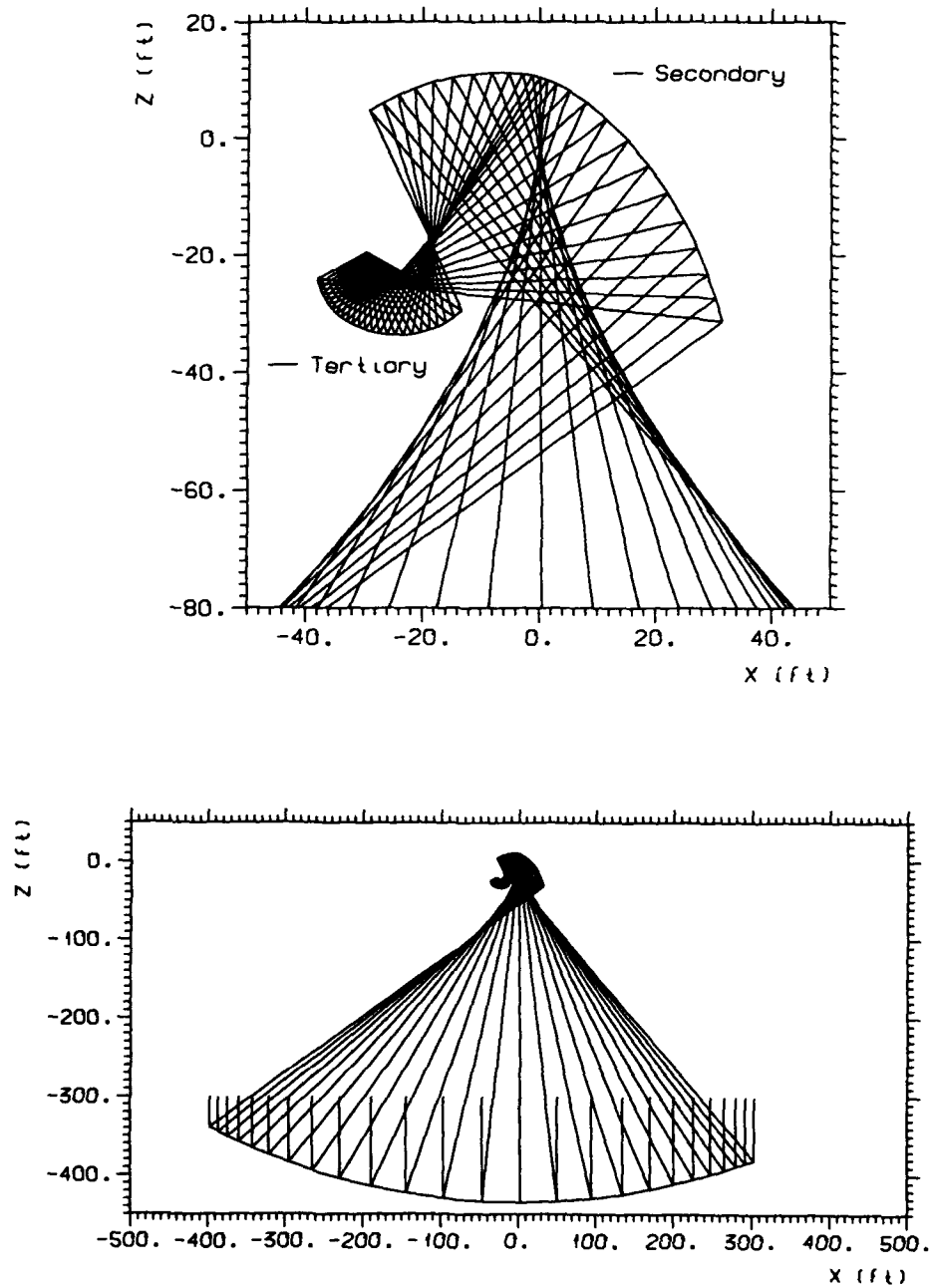
Figure 2   Cross-section of the final Gregorian dual-reflector feed and its ray paths in the symmetry plane. a) Dual-reflector feed. b) Complete tri-reflector system.
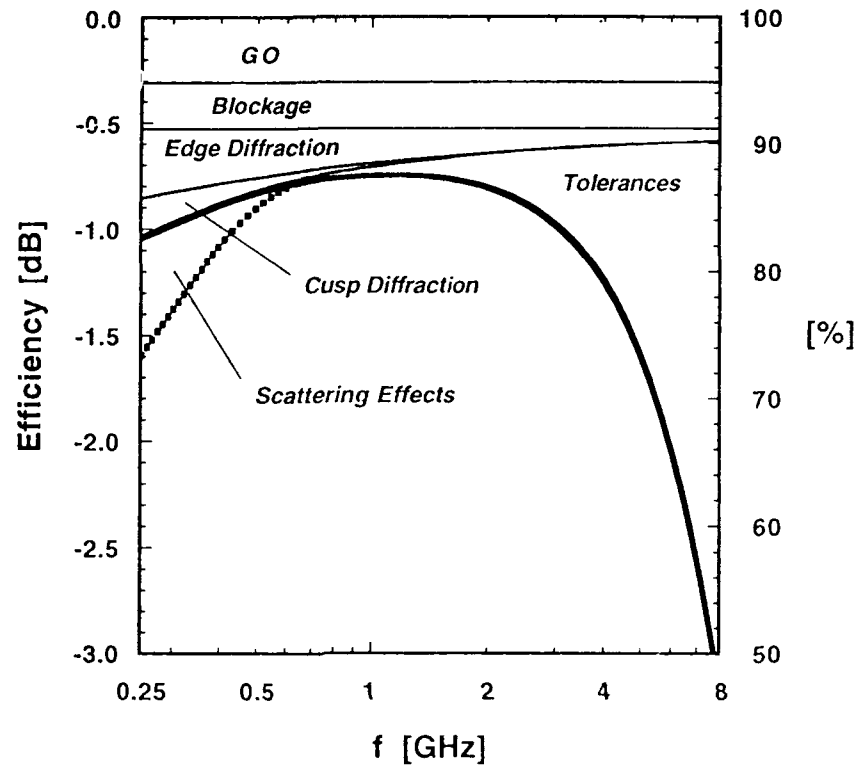
Figure 3   Different contributions to the total aperture efficiency.

ADVANCED MICROWAVE TECHNOLOGY IN MODERN COMMUNICATION SATELLITES

S. Jerry Fiedziuszko*

Introduction.

A rapid increase in global demand for voice, video, and data services has created explosive growth in capabilities of commercial communication satellites. A typical space system is shown in Figure 1.
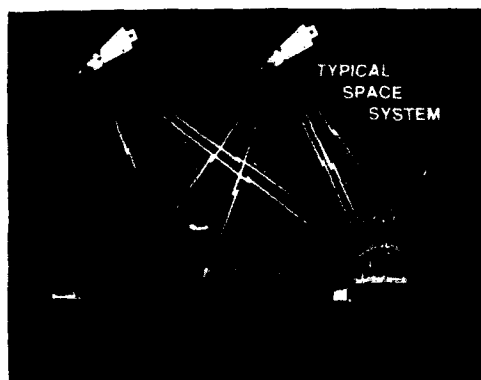
Figure 1. A typical space system

Since present communication satellites function as microwave relay stations and utilize a large number of sophisticated microwave components and subsystems, this poses a significant challenge to microwave engineers. To meet this challenge a large variety of novel microwave components have been developed.

Communications satellite systems can be divided into two categories: geostationary satellites (GEO) , which are relatively large with many transponders on board (Figure 2a), and low Earth orbit satellites (LEO), which are relatively small and used in constellations (Figure 2b). LEO satellite systems generally will be used in upcoming Personal Communication Systems (PCS).

* Space Systems/LORAL, Palo Alto, California, USA
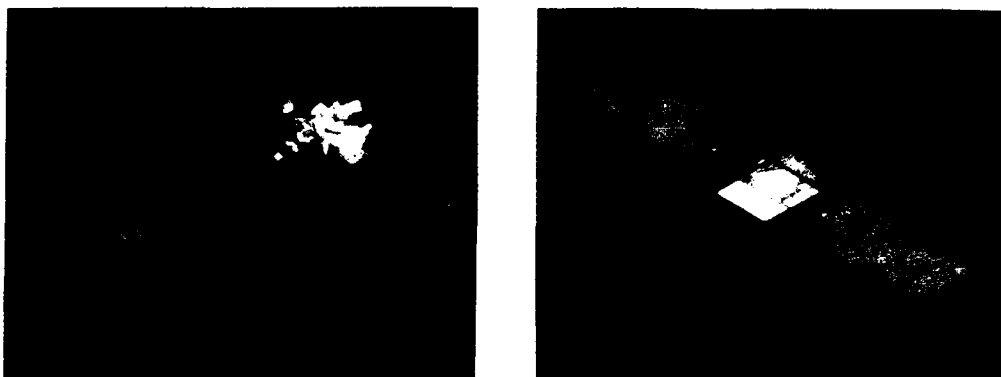
Figure 2. A communications satellite a) GEO satellite B) LEO satellite

Due to the limited size of the satellites and the need to have more communication channels available, reductions in size and weight of traditional microwave components become an absolute necessity. Key microwave innovations which have contributed to this increase in the number of transponders required to satisfy growing traffic demand include: miniature filters, MMICs, DROs, and phased arrays. A simplified block diagram of a typical geostationary satellite transponder is shown in Figure 3.
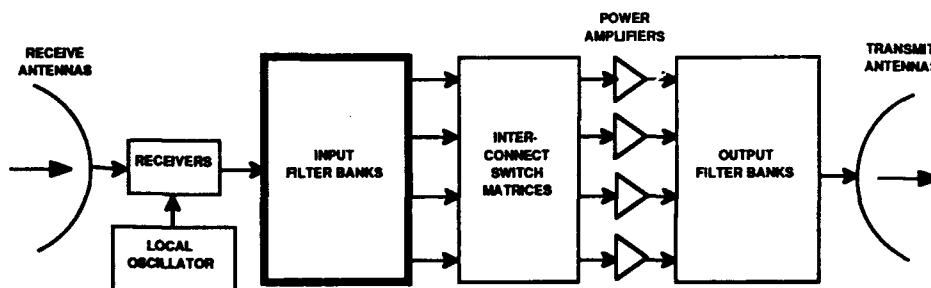
Figure 3. A typical GEO satellite transponder (simplified)

In the evolution of microwave circuitry in communication satellites initial systems utilized tubes, waveguides, a waveguide cavity or combline filters and individual components. The next step in this evolution was to increase the use of

microstrip and MIC implementations followed by a progressive trend toward increased integration. Advanced filter technologies were gradually inserted in satellite payloads (transponders). This included dual mode cavity filters, utilization of advanced composites (graphite) to reduce weight, contiguous multiplexing, and finally the use of dielectric resonators in single and dual mode operation. Recently, solid state power amplifiers have replaced TWT amplifiers @ 4 GHz (planned for 12 GHz), and hybrid technology for increased frequency bandwidth is being used. MMIC insertions were accelerated and the use of dielectric resonators in oscillators (DROs) was significantly advanced.

A typical cost and weight distribution in the communication payload of a GEO satellite is presented in Table I.

TABLE I.

Communication Subsystem Typical Cost and Weight Distributions

	Weight kg	Percentage Total Weight	Percentage Total Cost
Antenna	110.0	26	15
Multiplexers	115.0	28	30
Low Level Active Units	66.6	17	26
Power Amplifiers	86.7	21	25
Miscellaneous	31.7	8	4
Total Communication Subsystem	416kg	100% .	100% ($25M-35M)

Clearly demonstrated is the significant value of microwave components in a satellite (30-50%). Because of this large cost, the development of microwave components, which in the past was largely performance and reliability driven, now is increasingly focused on cost, size, and weight reductions. Average on orbit weight savings are in order of $50K/kg while power savings are approximately $8K/watt. In some cases savings in insertion loss for high power subsystems can be as high as $600k/0.1 dB. Therefore, increasingly, the size, weight and cost are being leveraged to arrive at an ultimate performance value increase for new circuit insertions. More and more the effort concentrates on a very careful total systems approach for an optimum solution considering the cost, producibility, and performance.

Monolithic Microwave Integrated Circuits (MMIC).
In the area of active microwave components this leads to:
•commonality of circuits•standardized circuits-standard carrier sizes
•modular approach •design with lower touch labor•simpler designs
•lower parts counts • decreased number of wire bonds•increased integration with MMIC's. As an example, MMIC Channel amplifier (shown in Figure 4) insertion savings are shown in Table II.
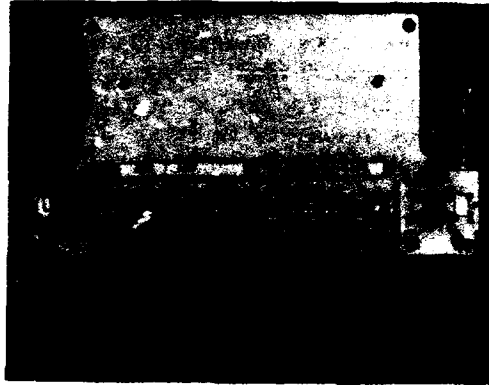
Figure 4. MMIC Channel Amplifier

TABLE II.

Weight/Cost Savings & Net Percentage Value Increase

Current Weight	940g/unit
New Weight	615g/unit
Weight Savings Per Flight (34/Flight)	11.1 kg
Weight Equivalent Savings	$685k
Value Added	$1.2M
Percentage Value	40%

Dielectric Resonator Oscillator (DRO).

In a typical communications satellite payload, frequency sources are one of the most expensive and labor intensive subsystems. Past communication subsystem frequency sources generally have started with a low frequency crystal controlled oscillator (TCXO), and used multipliers and amplifiers to reach required output frequencies.

Modern communication satellites use a low noise dielectric oscillator (oscillating at the required output frequency), which is phase locked (by use of a sampling phase detector or digital dividers) to a stable reference . A typical satellite DRO is shown in Figure 5. The use of GaAs and Si MMIC divider, scaler, and phase detector circuits in these designs significantly reduces overall complexity and labor cost.

In summary, the DRO has • lower parts count • higher reliability • less labor intensive • lower cost • smaller and lighter. Typical DRO benefits for GEO type satellites like Intelsat VII are listed in Table III.
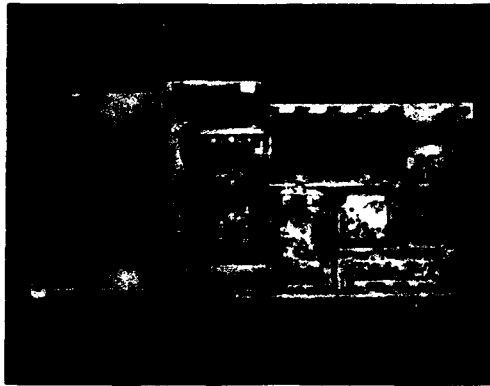
Figure 5. A typical satellite
DRO

TABLE III

• **Significant Benefits for The Typical Satellite
(Intelsat VII Type Satellite Savings)**

• Lighter Weight	4.3kg less per Flight
• Smaller Size	>20% Reduction in Volume
• Less Tuning	134 Fewer Circuits that Require Tuning
• Lower Cost	$800k Savings in Recurring Costs per Flight

Filters and Multiplexers.

A microwave signal received by a satellite, after amplification and downconversion by the receiver, is divided by the input multiplexer into individual channels for processing and power amplification. The high power signals are then recombined by the output multiplexer for transmission back to earth. Thus, the input multiplexer typically has one input and N outputs, while the output multiplexer has N inputs and one output. System specifications require these filters to be very narrowband, highly selective devices with demanding requirements for insertion loss, group delay flatness, return loss, bandwidth, rejection and insertion loss flatness. These demanding performance requirements dictate the use of highly sophisticated filter structures capable of elliptic function, group delay equalized responses. Current state-of-the-art technology uses dual mode cavity and dielectric resonator filters for satellite multiplexers. Some of them are shown in Figures 6 and 7.
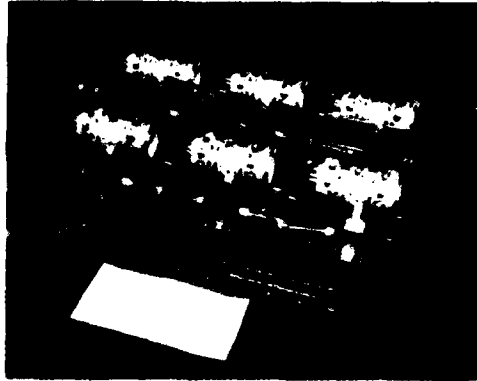
Figure 6. A satellite input multiplexer.
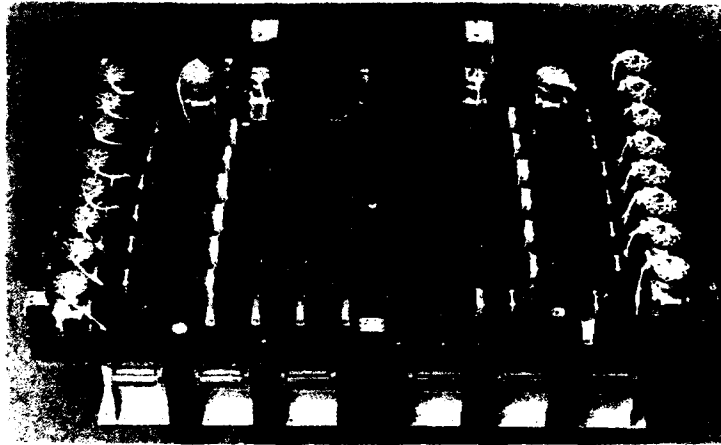
Figure 7. A satellite output multiplexer.

The output multiplexers (often contiguous) are typically fabricated of thin-wall INVAR and graphite epoxy to reduce weight while providing temperature stability, and the input unit, using dielectric resonators, represents the present-state-of-the-art. A typical GEO satellite contains several of these multiplexers and sometimes more than 100 filters. Their production is very labor intensive and their drawbacks are: an inherently large size (even with dielectric resonators), high mass and high production costs. The size and basic design results from the need to overcome the effects of conductive losses. This has created significant interest in High Temperature Superconductivity (HTS) for these applications. The evolution of satellite filters is presented in Figure 8, showing the potential impact of HTS filters and multiplexers.
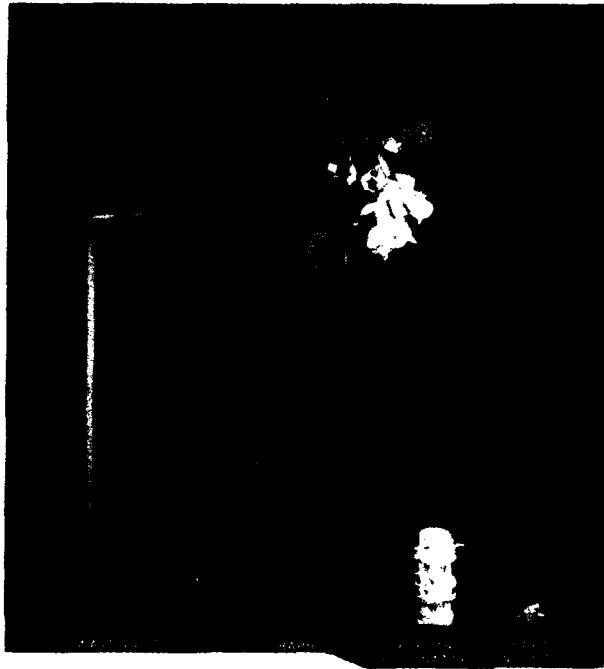
Figure 8. Evolution of satellite filters.

Phased Array Antennas for LEO Satellites.

The satellite industry is developing commercial space systems such as Iridium, Globalstar, and Ellipso to provide mobile communications to personal telephones (PCS). A typical LEO satellite mobile communications system is shown in Figure 9 [1].
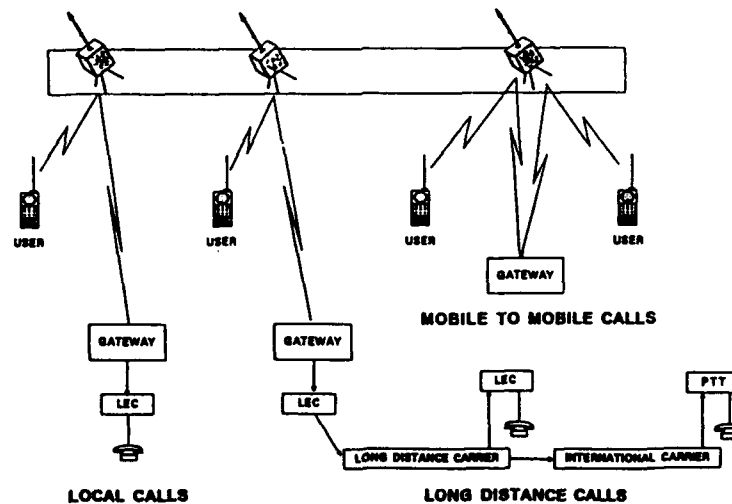
Figure 9. A typical LEO satellite mobile communications system.

One of the key subsystems in these constellations of LEO satellites is the phased array antenna. This multi-element antenna operates at relatively low microwave frequencies (1.6 GHz and 2.5 GHz), and contains many microwave components such as : radiating elements, filters, MMIC amplifiers, and phase shifters and combining/dividing networks. One of the possible configurations of such an antenna is shown in Figure10.

INTEGRATED ANTENNA ASSEMBLY   RF COMBINER LAYER   MMIC LAYER (BACK OF BOARD)   CALIBRATION SWITCH LAYER
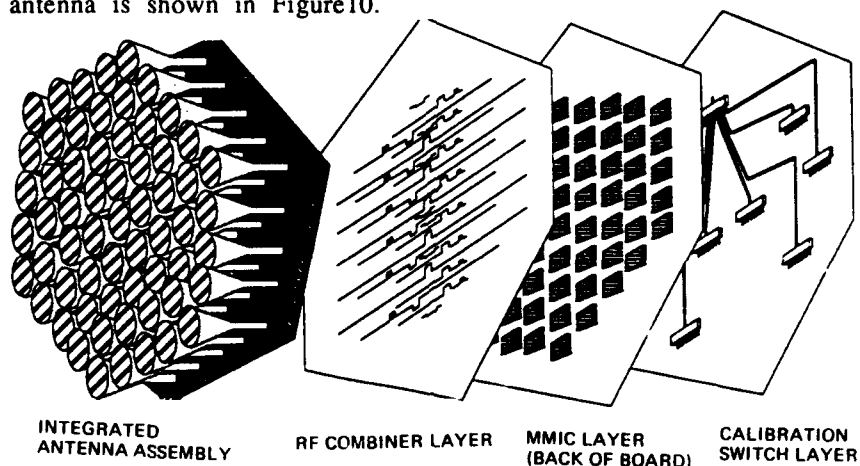
Figure 10. A satellite phased array antenna system.

At the present time significant efforts are being devoted to development of these components with low cost as the main objective due to the large number of components, as well as the large number of satellites required in viable LEO systems (e.g. Globalstar system requires 48 satellites).

Conclusions

In conclusion, satellite microwave technologies will remain highly cost driven. Due to the large number of high performance microwave components required in a typical communications satellite system, a higher use of modular , universal designs is needed. The use of MMIC's and DRO's is increasing very rapidly. Passive components such as filters and multiplexers require further miniaturization and High Temperature Superconductivity is considered as an option to reach this goal[2]. Phased array antennas combine antenna and traditional microwave technologies and require a large number of various microwave components.

References:

[1]. C-H. A. Tsao, P. Jung, G. Raguenet "An Isoflux Antenna for a Low Earth Orbit Satellite Mobile Communication System" AIAA, The 14th International Communication Satellite Systems Conference Digest, pp.639-649, Washington, DC, March 1992.

[2]. W. Morgan " Potential Uses of Superconductivity in Communications Satellites" AIAA, The 15th International Communication Satellite Systems Conference Digest, pp.609-614, San Diego, CA, March 1994.

DIFFERENTIAL METHODS OF SIGNAL SELECTION IN
MICROWAVE POLARIMETRY

ACAD,PROF,D-R Sc. KOZLOV A.I. - MOSCOW STATE TECHNICAL
UNIVERSITY OF CIVIL AVIATION, RUSSIA

The problem of detecting radar targets against the background of interfering reflections has been and will always be one of the most crucial tasks of radiolocation. In most of the radar stations the solution of the problem is reduced to comparing the output signal $U(t)$ to some threshold value - U_0. The rule of solving the task in this case is very simple, if $U > U_0$, then the target begin detected can be observed, if $U \leqslant U_0$, the target can't be observed.

In classical radiolocation, as a rule, the output signal $U(t)$ in proportional to evther module $|S_{11}|$, or squared module $|S_{11}|^2$ of one of the diagonal elements of the scatterilg matrix S, hence the information on the target and the background contai- ned in the remaining elements of the scattering matrix S, is irrevocably lost and is not used. Further applying two - channel technique where signals proportional to two diagonal elements of this matrix - S and S are the object for measuring open great additional possibilities for solving the problem of detection and result in socalled differential methods of differentiating between radar targets. In the latter, either equation $|S_{11} + AS_{22}|^n$, or equation $|S_{11}|^n + A|S_{22}|^n$ where $n = 1,2$, serve as a measured value, and coefficient A is chosen so that to ensure maximum contrast between the target being detected and the background. Note that the case where $A = 0$ will correspond to classical radiolocation.

Differentiation on the algorithms suggested above is convenient, first of all, due to the relative simplicity of their practical implementation with existing radars.

KOZLOV A.I, Professor, Doctor of Physical & Mathematical Sc.

Summary data relating to the dependence of the correct solution probability P_1 on the false solution P_2 with different algorithms of processing the signal being receivedare given in Table 1. Relay's Gaussian and exponential laws were chosen as underlying laws of the probability distribution. Expessions P_2 in lines 1 and 2 imply such a choice of value A at which value P_1 is maximum.

Table 1.

Correct Solution Probability P_1

N	Algorithms	Model	P	Figure
1 .	$|S_{11}| + A|S_{22}|$	Gauss	$1 - \Phi(q-kmb)$	1
1.1	$|S_{11}| - |S_{22}|$	Gauss	$1 - \Phi(q-kmb/\sqrt{1+h}\ )$	2
2 .	$|S_{11} + AS_{22}|$	Raley	$P_0\ 1/1+m^2\gamma^2$	3,4
2.1.	$|S_{11} - S_{22}|$	Raley	$P_0\ 1/1+m^2\gamma^2$	5
3.	$|S_{11}|^2 + A|S_{22}|^2$	Exp.	$P_0\ 1/1+m^2 f(\gamma)$	6
3.1.	$|S_{11}|^2 - |S_{22}|^2$	Exp.	$P_0\ 1/1+m^2 f(\gamma_1)$	

Parameter q is determined by value P_2 in the following way: at $P_2 = 10^{-7}$, $q = 5,0$, at $P_2 = 10^{-6}$ $q = 4,7$, at $P_2 = 10^{-5}$ $q = 4,3$, at $P_2 = 10^{-4}$ $q = 3,7$. Coefficient K shows the ratio between square of mathematical expectation and dispersion of value $|S_{11}|$ for the phones object. Parameter m is equal to the ratio of the squares of values $|S_{11}|$ for the cases of the presence and absence of the being detected. The equation for parameters actually determining the expediency of differential methods of differentiation are determined by equation

$$b= \frac{1}{\sqrt{1-\rho_0^2}} \sqrt{1+\alpha^2/c^2-2\rho_0\alpha/c} \ , \qquad (1)$$

where ρ_0 – *correlation coefficient for a signal in the absence of the target, a and c – sizes of the assymetry of the scattering matrices in the cases of the presence and absence of the target, respectiely. (The size of assymetry implies the ratio of mean square values of modules* $|S_{22}|$ *and* $|S_{11}|$*). The case corresponding to classical radiolocation occurs at* $b = 1$.

The characteristics of differentiation of radar targets corresponding to the first algorythm of the received signals processing are presented in Fig.1.

A particular case of the first algorythms in given in line 1.1. It corresponds to $A = 1$*. Parameter h in the equation is connected with the non-optimum choice of coefficient A. The parameter itself may be found throngh the following equation:*

$$h = m^2 \frac{1+a^2-2\rho_1 a}{1+c^2-2\rho_0 c} \qquad (2)$$

where ρ_1 – *correlation coefficient of the received signal in the case the larget is available.*

Parameter b in the same equation is equal to:

$$h_1 = \frac{a-1}{\sqrt{1+c^2-2\rho_0 c}} \qquad (3)$$

The efficiency of using the algorythm 1.1 may be assessed by comparing the values b_1^2 *and* $(1+h)$*. As an illustration, the curves corresponding to* $\rho_1 = 0,9$*;* $\rho_0 = 0,85$*;* $m = 0,5$ *are given in Fig.2.*

Algorythm 2 given in line 2 of Table 1 results in a different equation for probability P_1*. Due to the cumbersomeness of the dependence of* γ *on parameters a, c,* ρ_0*,* ρ_1 *its evident type is not given. (For classical radiolocation* $\gamma = 1$*). As an illustration, the dependence of* γ *on the ratio a/c for some particular targets and phones is given in Fig.3, and corresponding characteristics of differentiation Fig.4.*

A particular case for the algorythm under consideration in given in line 2 where the designation $\gamma_1 = (1-\rho_1)/(1-\rho_0)$ *is introduced. The efficiency of this method is illustrated in Fig.5.*

For algorythm 3 and its particular case 3.1 the evident
type for ρ_1 in Table 1 is not given because of the excessive
cumbersomeness of functions $f(\gamma)$ and $f(\gamma_1)$. For obriousness, in
Fig.6 the dependence of P_1 for algorythm 3 on   , corresponding
to classical radiolocation $(f(\gamma)=1)$ is givenm where it is assumed
that $\rho_1 = 0.85$.

In conclusion some results of experimental measurements ob-
tained by the authors in using algorythms 1.1 and 2.1 are given.

The earth surfacc artificially covered with brick, planks,
turf, slate and sand in a chaotic manner was chosen as an object
for observation. A small metal rod was placed on that surface.
(The situation typical for the experiment is given in Fig.7). In
the process of the displacement of the experimental installation
(Fig.8) it was required to detect the metal rod. In the course
of the experiment $|S_{11}|$, $|S_{22}|$, $|S_{11}|-|S_{22}|$, $|S_{11}-S_{22}|$ were deter-
mined simultanevusly. Then their statistical processing was car-
ried out. Only some data testifying to the availability of great
possibilities for differential methods of differentiating radar
targets will be given here. As a result of 572 measurements of
the square at different angles the following data were obtained
for the case given in Fig.7: in the absence of the target $|S_{11}|=$
$= 180 \pm 10$; $\sigma_{11} = 80 \pm 2$; $|\overline{S_{22}}| = 185 \pm 10$; $\sigma_{22} = 81 \pm 2$, in the
presence of the target $1S$ $1 = 190 - 10$; s $= 72 - 2$; $1S$ $1 =$
$= 186 \pm 10$; $\sigma_{22} = 81 \pm 2$. As it can be seen, any detection of
the target being detected is out of question here. There is qui-
te a different picture with the other two algorythms the mean
value of the difference $|S_{11}| - |S_{22}|$, in the absence of the tar-
get, amounted to 2.1 ± 0.3, and in its presence - 45 ± 3, the
mean-square value in the first case being 2.04 ± 0.02, and in the
second case - 11.0 ± 0.1. The same was the case in determining the
mean value of the module of the difference $|S_{11}-S_{22}|$, which
amounted to 2.1 ± 0.3 and 87 ± 5, respectively, with mean square va-
lues being 4.17 ± 0.03 and 22 ± 6, respectively. The result obtained
convincingly testify to great additional possibilities which can
be provided by differential methods of differentiating radar sig-
nals in detecting some classes of radar targets.
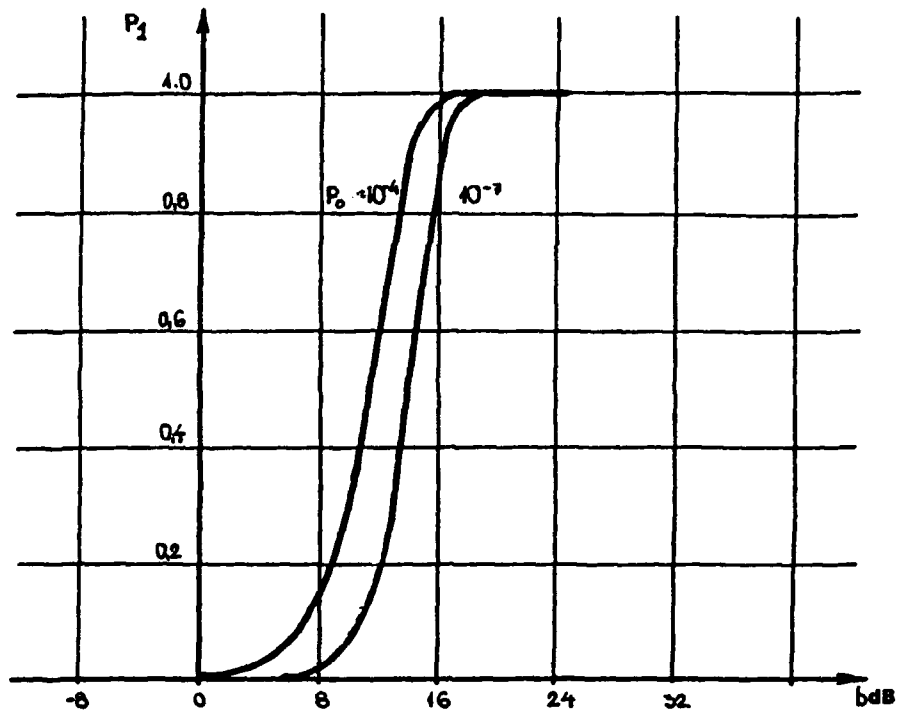
Fig. 1

Characteristics of differentiating radar targets (algorythm 1)
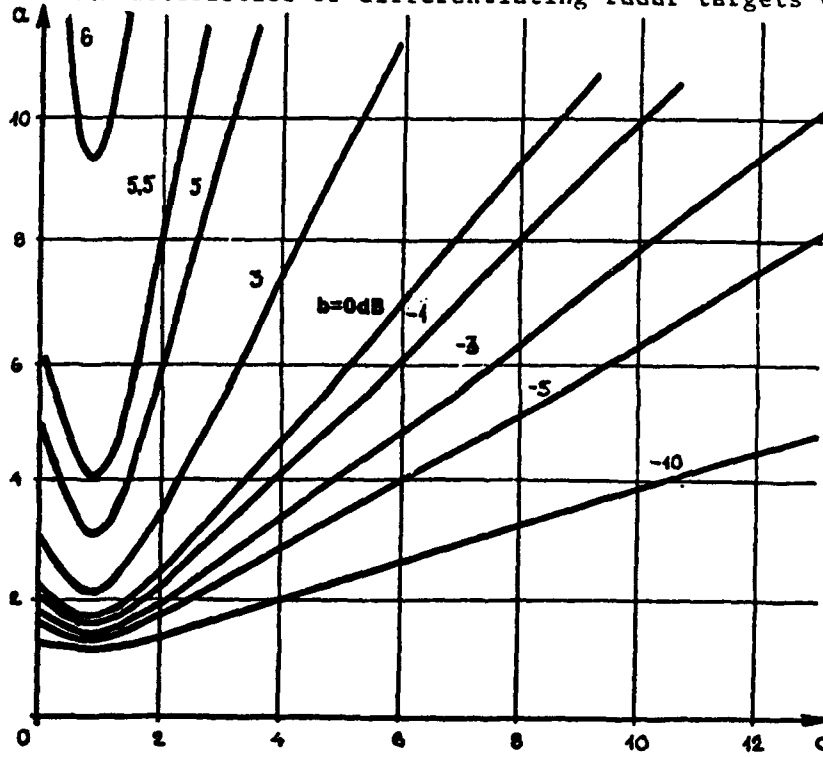
Fig. 2

Efficiency zones boundaries (algorythm 1.1)
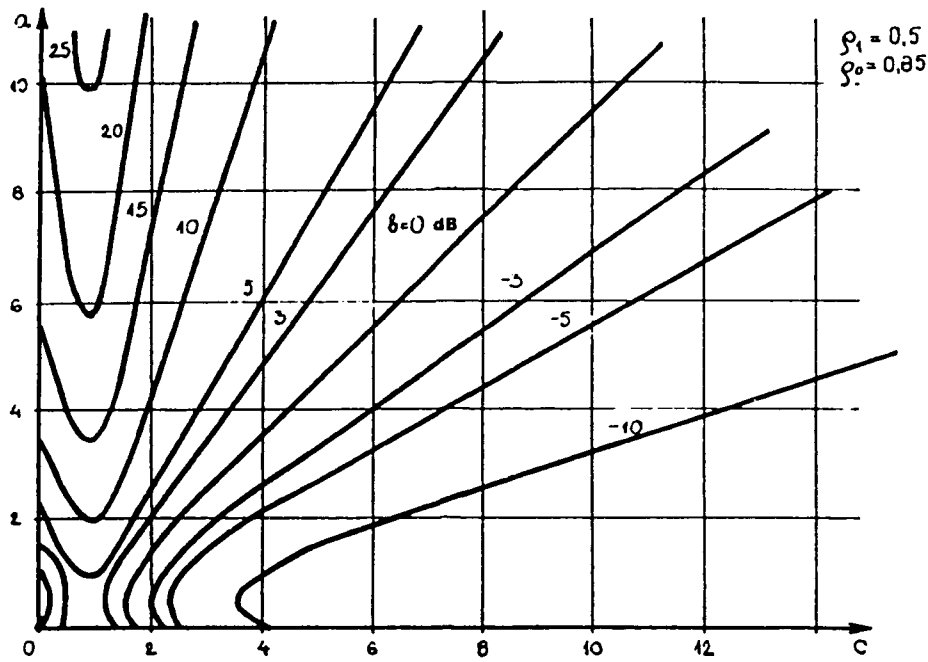
Fig. 5

Efficiency zones boundaries (algorytm 2.1)
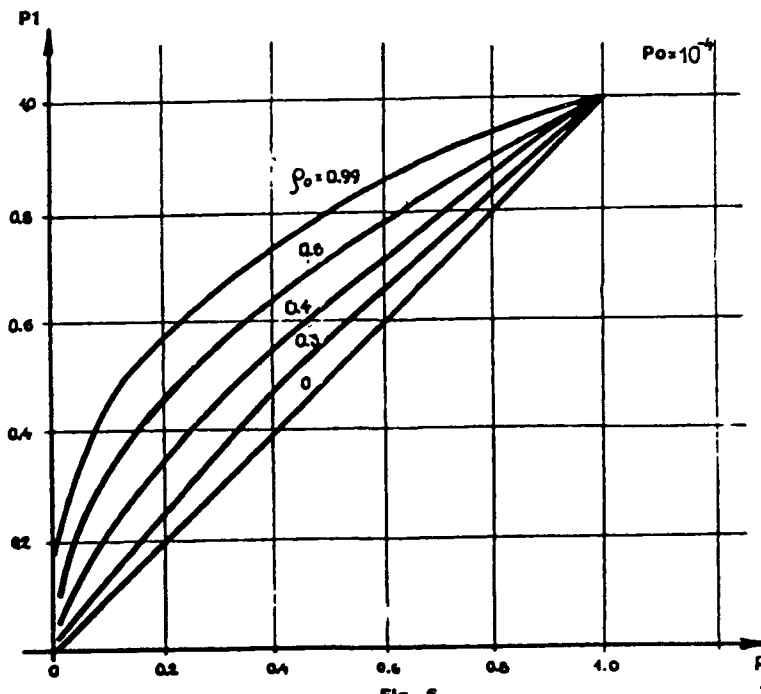
Fig. 6

Dependence of P1 on P2 (algorythm 3)
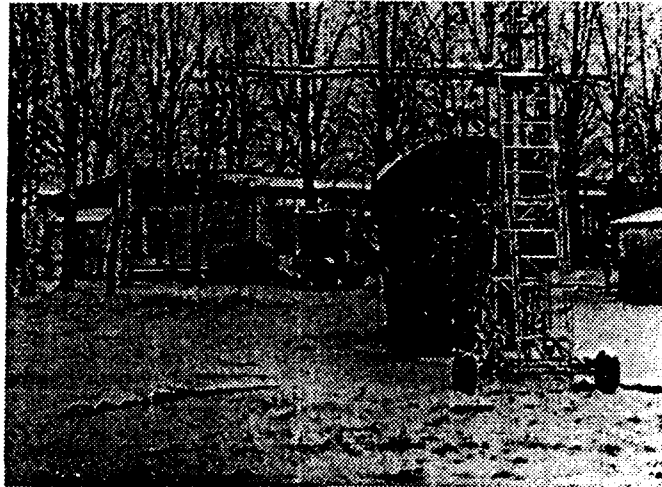
Fig. 7

General view of the experimental installation
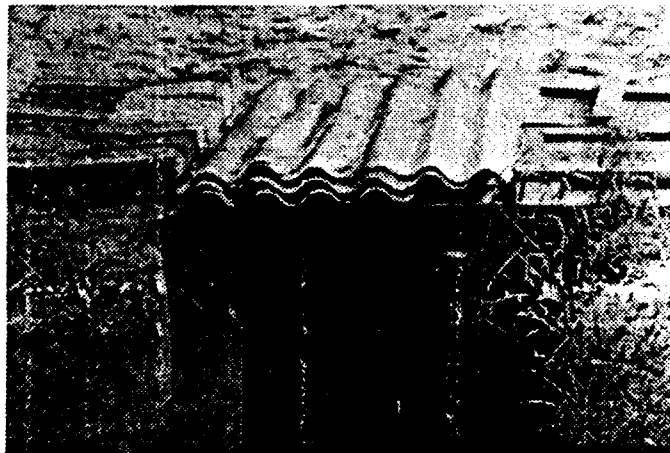
Fig. 8

Artificial surface under observation

ANTICOLLISION CAR RADAR IN THE mm-WAVE RANGE WITH PSEUDO-NOISE-CODE MODULATION AND DIGITAL ANGLE EVALUATION.

by   H.P. Groll, J. Detlefsen, M. Rozmann, T. Troll, *

1. Introduction and System Concept

Former anticollision car radar systems have been under investigation since many years but the experiments have shown, that the false alarm rate in these systems mostly was too high. One of the reasons was the lack of an angular resolution. It seems to be necessary for a car system also to find out the lateral position of a target. A new collision avoidance system is under investigation since five years at the Technische Universität München within the European PROMETHEUS project. The choice of the modulation method depends mainly on the application. It must be suitable for a multi-target situation and resistant to interferences between a great number of cars, which can be equipped also with such radars. A radar with Pseudo-Noise (PN) code modulation can be changed in 4 different properties: The code, the code-repetition frequency, the carrier frequency and possibly also the polarization. To measure the relative velocity of cars the Doppler frequency must be evaluated, which needs a coherent system. For the angular allocation an antenna array with 1 transmitting antenna and 4 receiving antennas is used, where the phase discrimination is made between the 4 receiving channels in a digital way. This seems to be cheaper than a conventional

Fig.1   System Block Diagram

---

* Technische Universität München, Germany, Microwave Department

phased array. The necessary field of view according the usual dimensions of motor-ways is characterized for our experimental system in a horizontal opening angle of 12°. In a distance of 75 m three lanes can be covered and in a distance of 100 m the field of view corresponds to five lanes. The vertical angle of only 3° avoids reflections from bridges and tunnels. In the first step a separation of 4 angular resolution cells is used, but later an improvement to 8 angular cells, using a switched second transmit-ting antenna [4], is planned. In radial direction we have 128 resolution cells of 75 cm each, which means a whole distance of 96 m. For a range of 100 m the first cell starts at a distance of 4 m.

2. Range Resolution

The PN-code has a clockrate of 200 MHz and a code length of 1023 chips. The code repetition frequency is 196 kHz. So the unambiguous range is 767 m. The width of one bit of the code is 5 ns. Ac-cording Fig.1 , which shows the block diagram of the whole setup, the binary phase modulation is applied to a subcar-rier of 1,2 GHz. Then the signal is upcon-verted to 61 GHz. The upper sideband is selected by a waveguide filter. The carrier frequency of 61 GHz has been allocated by the German PTT. (In the future the frequency will be in Europe probably 77 GHz). Other carrier frequencies can be used easily by changing only the mm - wave components. The maximal - length pseudo noise code is not generated by shift registers, as usual, but read out from a memory. All codes, both for the trans-mitted signal and for the 1023 reference codes, which are shifted by one bit each, are stored in EPROM's. The ac-cess to a desired range gate
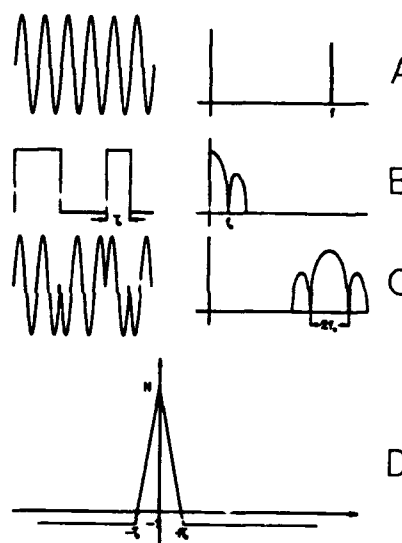
Fig.2 BPSK-Modulation
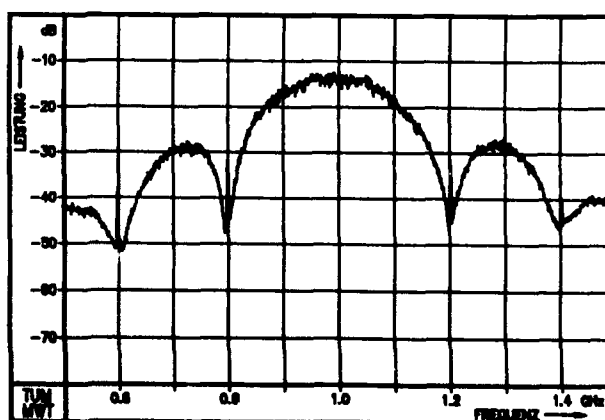
Fig.3 Measured Spectrum

can be made quite easily via a code shift by simple address selection. Processing all 128 range gates at a rate of 10^4 range cells per second, the wholefield of view can be covered within 13 ms. Fig. 2 shows the principle of the modulation, where A is the carrier, B the PN code, C the modulated signal and D the autocorrelation function. Fig. 3 is a measured spectrum of a modulated signal. The main lobe is 400 MHz wide. The theoretical gain of the system is 60 dB. It is possible to cover a distance of 150 m with a single sideband power of only 1.6 mW. A measured crosscorrelation signal of a single point target is shown in Fig. 4. It indicates, that near range side lobes are at 40 to 45 dB,

Fig.4 Crosscorrelation Signal

the average signal to noise ratio is better than 50 dB.
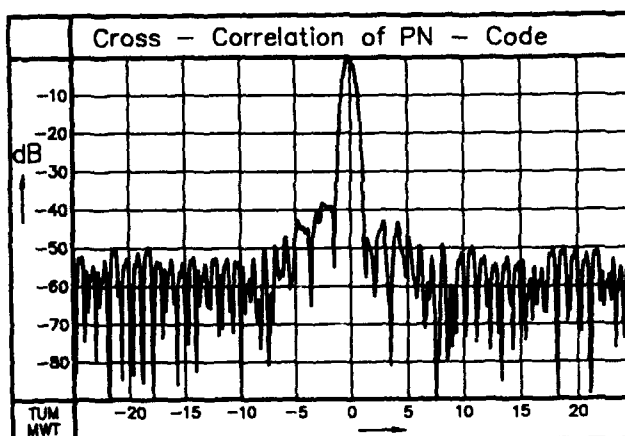
3. Lateral Resolution

The mm-wave front end is explained in Fig.5. It consists of two injection locked Gunn oscillators to provide the carrier and the power for the mixers. As shown in Fig.1 and Fig.5, the signals of the four receiving antennas are downconverted with the transmitter frequency by four homodyne quadrature mixers, which produce I - and Q - signals in the baseband. These eight signals are correlated in the receivers with the reference code, which is shifted in time according the wanted range gate. After ana-
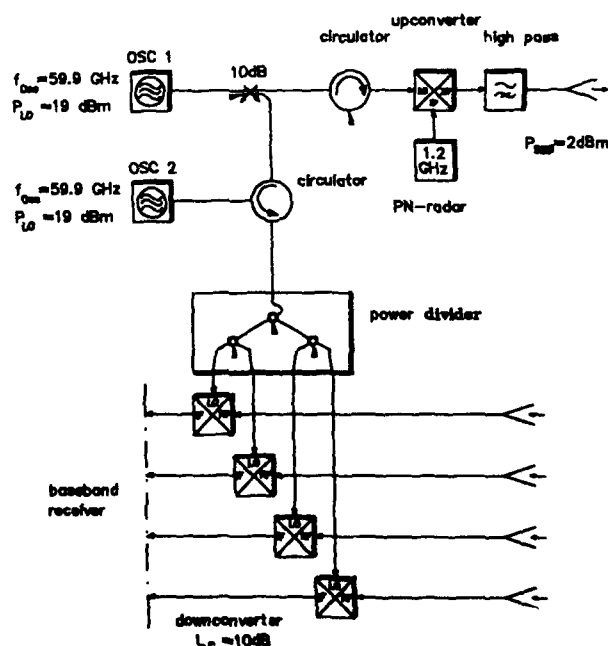
Fig.5 mm-Wave Front End

log to digital conversion a signal processor generates calculated beams for angular discrimination by the so called wavefront reconstruction [3,4]: Four suitably weighted samples of the complex field distribution received by the four antennas are transformed by a FFT into the response of an equal number of antenna beams of ca. 3° beamwidth. The eva-luation of the angle of an incident field of the wave-front can be explai-ned by the phase comparison of the voltages of the four channels according the drawing of Fig. 6. The field distribu-tion of the incoming wavefronts, which is sampled by the receivers, can alternatively be evaluated by the more sophisticated methods of spectral estimation. Depending on the signal to noise ratio this can increase the effective angular resolution beyond the classical limits. Because the reconstruction is carried out by digital signal processing, the number of resolution cells and the area to be surveyed can be adjusted automatically according to distance and depending on traffic requirements.

Voltage/Time   Reflected Wavefront

4 Receivers

Ant 1
Ant 2
Ant 3
Ant 4

α

$$\alpha = \arcsin\left(\frac{\Delta\varphi \cdot \lambda}{2\pi \cdot \Delta x}\right)$$

Δx : Aperturbreite
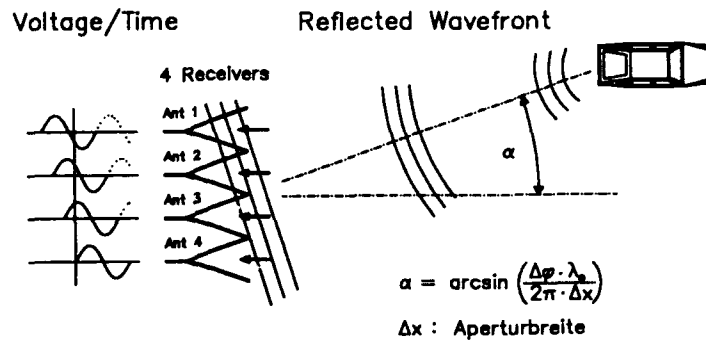
Fig.6   Wavefront Reconstruction

4. Velocity Measurement

In case of an evaluation of the phase of the autocorrelation function at two consecu-tive times with a sampling rate of 50 kHz and a system bandwidth of 20 kHz the velo-city resolution will be 14 km/h and the velocity ambiguity ±177 km/h in this case. If one or more objects are detected in one spatial resolution cell, one can try a velocity estimation by evaluating a small number of values using the phase autocorrelation method. If the rela-tive movement of the object is critical, it can be tracked by applying a FFT.

5 . System Evaluation

After construction and testing of the four chanel system in the laboratory it was mounted in an experimental vehicle. The mm-wave frontend is made in conventional waveguide technique and for the first step the antennas consist now of rectangular horns with cylindrical dielectric lenses. The aperture of one horn is 21 mm * 70 mm, the whole width of 4 horns is 88 mm. The depths of a horn is 60 mm, shown in the

drawing of Fig. 7. The vertical beamwidth is about 2.5°, the horizontal beamwidth of one antenna is ca. 11.5°, the angular resolution of the 4 antenna combination is 3°. The sidelobes are below 25 dB. The mm-wave setup, which is shown together with the 5 antennas on the photograph of Fig. 8, was mounted on the front side of the car and is covered by a thin dielectric radome. Together with the radar setup a video camera and facilities for data recording are provided in the car. The purpose is to make simultaneous recordings of typical sequences of radar data and video pictures and their superposition. So one can optimize the data processing routines in the laboratory using real radar signals, which were measured in realistic traffic conditions. The radar data are stored in digital form on a hard disk. All the video pictures have a number and can be compared easily with the data which belong to the measuring time of the video image, which has a repetition frequency of 15 Hz, and the four received radar signals, which are displayed on a monitor in the car like usually at an A-scope. This gives the possibility to judge the quality of the radar signals during driving. In the first experiments the traces of the radar signals were written on the
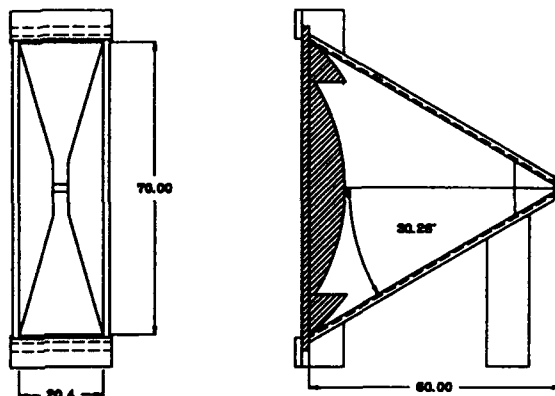
Fig.7 Horn Antenna with Dielectric Lens

Fig.8 mm-Wave Frontend with 5 Antennas

display from left to right - like in the A-scope - and the 4 traces coming from the 4 antennas were one above the other. The lowest trace corresponds to the right an-

tenna and the upper trace belongs to the left receiving channel. This kind of picture is shown in the photograph of Fig. 9. In the same way - but horizontally expanded are the raar traces on the picture of Fig. 10. This corresponds to a kind of focusing procedure, which was made in the first time at the beginning of a drive: By positioning a corner reflector in the center of the front of the experimental car one had to adjust the amplification and the phase of the four receiving channels. The corner reflector and the same size of the four reflected signals one can see
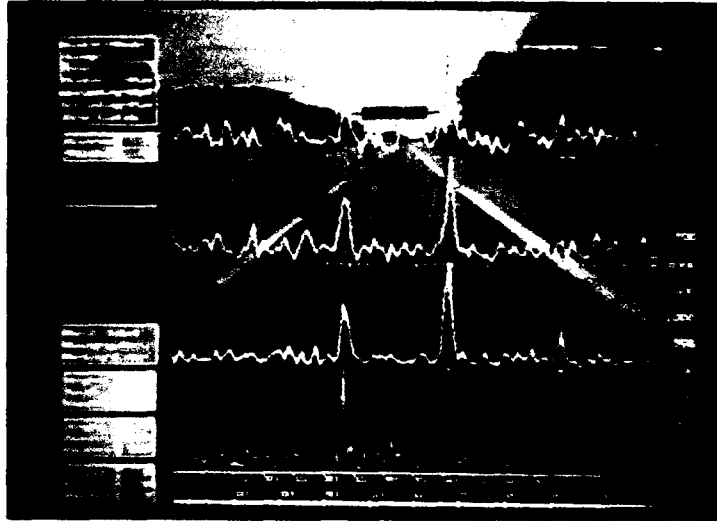
Fig. 9   Video picture with 4 radar traces
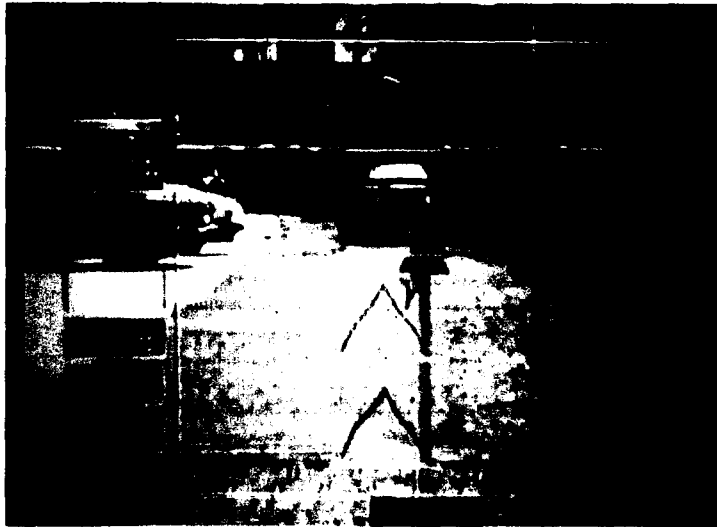
Fig. 10   Expanded radar signals during focusing

on the photograph of Fig. 10. Later the traces were turned 90° and written on the screen from below to the top. So the optical impression of the picture is near to the

real geometry. The Fig. 11 shows a photograph, which is made from a video tape

record after a drive
on a highway. The
left trace belongs
to the left recei-
ving channel, the
right trace to the
right antenna. The
distance scale on
the left side indi-
cates as maxi-
mum value 80 m,
which was chosen
for the first experi-
ments [6 to 13].
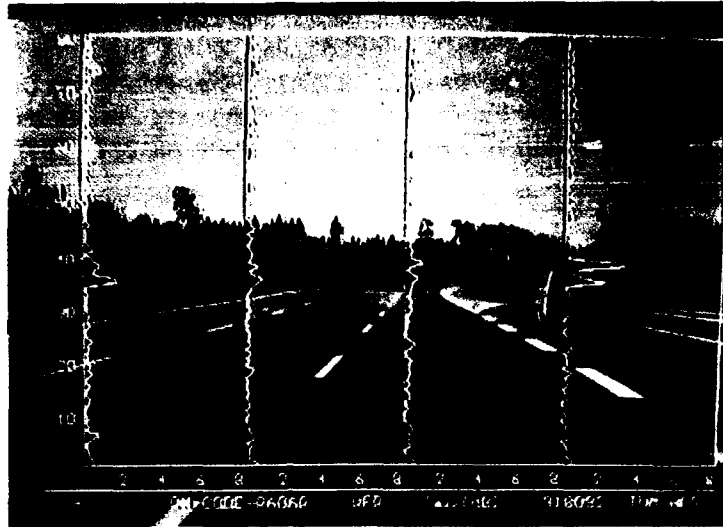
Fig. 11   Video picture with 4 vertical radar traces

6. Conclusions

The system mentioned above is not any kind of a prototype. The purpose of our investigations is to find out the technical possibilities like the necessary transmitted power, the sensitivity of the receivers, the suitable modulation system and the right way of signal processing, both for the distance resolution and for the angle evaluation. It is obvious, that a mm-wave front end in milled waveguide technique will be too expensive. The future of such systems can be solved only by introducing MMIC´s, which can be developed only in case of an application in a very high number of items. Also the antenna design will need a flat and space saving solution.

References

[1] Lindner,K.; Wiesbeck,W.: 35-GHz-Impulsradarsensor zur Verwendung in Ab-
standswarngeräten für Kraftfahrzeuge. Nachr. Techn. Z. 29 (1976), pp. 667.

[2] Neininger,G.: An FM/CW Radar with High Resolution in Range and Doppler App-
lication for Anti-Collision Radar for Vehicles. IEE Conf. Publ. No.155: Radar 77
London (Oct. 1977), pp. 526 - 530.

[3] Detlefsen,J.: Abbildung mit Mikrowellen. Fortschrittsberichte der VDI-Zeitschrif-
ten, Row 10, Volume 5, VDI-Verlag, Düsseldorf (1979).

[4] Bockmair,M.; Detlefsen,J.: A Multistatic Two-Dimensional Millimeter Wave Imaging System at 35 GHz. Proc. 13th Europ. Microwave Conf. Nürnberg (1983), pp. 540 - 545.

[5] Rozmann,M.; Detlefsen,J.; Lange,M.: Phasencodiertes Dauerstrichradar bei 1 GHz mit hoher Empfindlichkeit und Auflösung. 7th DGON - Radarsympos. Ulm (1989), Verlag TÜV Rheinland, Köln 1989, pp. 377 - 387.

[6] Detlefsen,J.; Rozmann,M.: PN-Code Millimeter-Wave Radar. Proc. 1990 West Germany MTT/AP Chapt. Workshop MICROW. COMPONENTS AND SUB-SYSTEMS; (Febr.8-9, 1990), Schloß Reisensburg.

[7] Rozmann,M.; Lange,M.; Detlefsen,J.: Collision Avoidance Radar Using Wavefront Reconstruction. Proc. 3rd PRO-CHIP Workshop, Paris (May 14th 1990), pp. 251 - 256.

[8] Zeilinger,W.; Zundl,T.; Detlefsen,J.: A Collision Warning Radar Using PN Code Ranging and Wavefront Reconstruction. Proc of the Workshop of 21st European Microwave Conference, Stuttgart (Sept. 13th, 1991), Microwave Exhibition and Publishers, London, pp. 153 - 158.

[9] Detlefsen,J.; Troll,T.; Zeilinger,W.: Collision avoidance radar using wavefront reconstruction. Proc. 6th Prometheus PRO CHIP Workshop, Kista, Sweden, (May 1992), pp. 113 - 115.

[10] Detlefsen,J.; Troll,T.; Rozmann,M.; Zeilinger,W.: System aspects and design of an automotive collision warning PN code radar using wavefront reconstruction. IEEE-MTT-S. Digest, Albuquerque, New Mexico, (June 1992), Vol.2, pp. 625 - 628.

[11] Detlefsen,J.; Rozmann,M.; Troll,T.; Zeilinger,W.: Auffahrwarnradar für Kraftfahrzeuge - Evaluierung des Systemkonzepts. Proc. Mikrowellen und Optronik Conf. (MIOP), Sindelfingen, Germany (May 1993).

[12] Troll,T.; Detlefsen,J.; Rozmann,M.; Zeilinger,W.: Dopplerfestigkeit eines phasencodierten Dauerstrichradars für KFZ-Anwendungen. Proc. 8th DGON-Radar-Symposium, München, Germany (Sept. 1993), Verlag TÜV-Rheinland, pp. 31 - 36.

[13] Groll,H.; Detlefsen,J.; Rozmann,M.; Troll,T.: Car Collision Avoidance Radar Using mm-Waves With PN-Code Modulation and Digital Wavefront Reconstruction. (Design and Experimental Results). Proc. 4th Internat. Sympos. on Recent Advances in Microvave Technology (ISRAMT´93), New Delhi, India, Dec. 15 - 18, 1993, pp. 735 - 738.

INDUSTRIAL MICROWAVE SENSORS

Ebbe Nyfors*, Pertti Vainikainen*

ABSTRACT

Microwave sensors are used for a variety of applications in the industry, in medicine, and for research purposes. This paper gives a review of the different types of sensors with classical and new examples. Special emphasis is given on sensors developed in Finland.

INTRODUCTION

The demand for sensors for the measurement of most diverse quantities has grately increased with the automatisation of industrial prosesses. In many cases microwave techniques provide competitive solutions. A large number of important applications are found in the field of moisture measurement, but many other material quantities as well as dimensions and movement are also measured using microwaves. The major advantages with microwave sensors are the capabilities to measure nondestructively, without contact from a short distance, using penetrating waves, and without health hazards to the personnel. Disadvantages are the usually high degree of specialization and the simultaneous existence of several variables effecting the microwave measurement (temperature, density, moisture, structure, etc.) in material measurements. There are, however, some sensor types, like for example the free-space transmission sensors that are used for the measurement of many different materials. The basic calibration against the quantity to be measured is done for each material separately. The problem with many variables can be mastered by using the multiparameter technique (resonant frequency & quality factor, phase & attenuation, several frequencies, or combination with other techniques) [1, pp.80-87].

RESONATOR SENSORS

A microwave resonator is made of a section of transmission line with open or shorted ends. Depending on the type of transmission line, the resonator is called for example coaxial, microstrip, stripline, slotline, or cavity resonator. When the resonator is used as a sensor, the object to be measured is brought into contact with at least some part of the electromagnetic field in the resonator. As a consequence, the resonant frequency and the quality factor will change in relation to the permittivity of the object. Because of the large variety of possible structures, sensors can be designed for measurement of almost any kind of object. Resonators can for example measure thin films, slabs, threads, surfaces, gases, liquids, powders, or granular materials. Resonator probes can be pushed into soft materials and for example an array of stripline resonators can measure the real-time moisture profile of a paper web in a paper machine, without touching the paper.

The Swedish company, Skandinaviska Processinstrument Ab (Scanpro), has developed a range of both hand-held and mechanically scanning cavity resonator sensors for the measurement

*Radio Laboratory, Helsinki University of Technology

of moisture in paper, pulp, felt, and cellulose. The first model came on market in 1968 and the development of the sensors has continued since then. Today their share of the market is substantial. The sensors for the on-line measurement of paper (Figure 1) are usually in the dry end of the paper machine. They are split cavities employing two resonant modes. One mode is less affected by the paper than the other mode, thus providing compensation for thermal expansion and humidity variations. The dry weight of the paper must be known for the calculation of the moisture because of the one-parameter nature of the paper measurement.
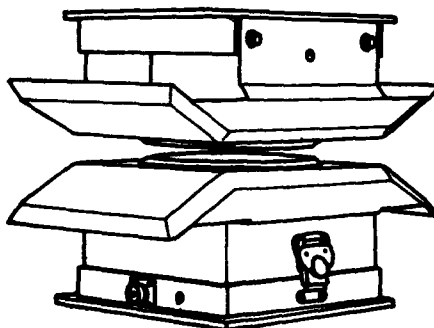
Figure 1. Split-cavity sensor made by Scanpro for the measurement of moisture in paper.

TRANSMISSION SENSORS

The basic transmission sensor consists of two horn antennas, a transmitter and a receiver. The object to be measured is put between the antennas in such a way that the microwaves pass through the object. The attenuation or the phase shift are measured. Such a free-space transmission sensor is suitable for measurements on conveyor belts or in large pipes (Figure 2). This type of sensor is the most often encountered microwave sensor in the industry. The major problem with free-space transmission sensors is caused by the multiple reflections inside the object and in the space between the antennas and the object. The difficulties are usually avoided by ensuring that the attenuation in the object is high enough (≥ 10 dB), by using oblique transmission, large distance between the object and the antennas (limited by diffraction), or frequency sweep. In some sensors the microwaves are guided from the transmitter to the receiver by a transmission line. Such guided wave transmission sensors are used for example for the measurement of liquids (e.g. microstrip structure) or films (e.g. split waveguide). They resemble resonator sensors, but are better suited for measurement of high-loss materials.
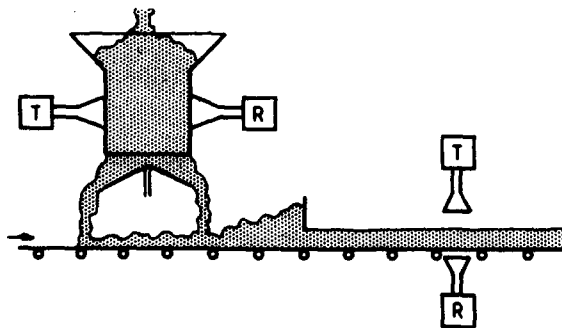
Figure 2. Free-space transmission sensor mounted for measurement on a conveyor belt or a pipe.

In Germany Laboratorium Prof. Dr. Berthold has developed a free-space transmission sensor for use on conveyors. Both the attenuation and phase shift are measured, and a radioactive sensor is used for the compensation of variations in the thickness of the material layer. The instrument is versatile, but an important application is the measurement of moisture in coal [2].

In Poland the Wiltech company has developed a versatile instrument [3]. The control unit can be used with both guided wave and free-space transmission sensors. Sophisticated modulation techniques are used to compensate for reflections and other sources of error.

In the U.S.A. the company Kay-Ray has developed a free-space transmission sensor mounted on a square pipe. Only the attenuation is measured but the result is combined with the result from a radioactive density sensor.

REFLECTION SENSORS

The reflection sensors are based on the measurement of the reflection coefficient, for example from the open end of a transmission line. A widely used sensor is the open-ended coaxial sensor. The open end is held against the surface of the object or is immersed or pushed into it. The fringing field at the tip penetrates the surface of the sample. The sensor is well suited for research purposes, in which case the advantages are the broad bandwidth (typically two decades), the small measurement area, and the ease of sample preparation.

Hewlett Packard offers a coaxial sensor to be used with a network analyzer in the laboratory. The frequency range is nominally from 200 MHz to 20 GHz, but the upper frequency limit is lower for samples with high permittivity. The software provided performs the measurement, calculates the results and automatically displays the real and imaginary part of the permittivity for the chosen frequency range. The sensor provides a fast and easy means to measure the permittivity, but only flat and homogeneous samples can be measured because of the flat sensor tip and small measurement area. Any air gap between the sensor and the sample (e.g. because of a rough surface) immediately lowers the measured values.

Reflection measurtements can also be accomplished in free space. For example at the Belorussian Academy of Sciences in Minsk, millimeter wave ellipsometric methods for the investigation of dielectric coating materials have been developed [4]. In ellipsometry the ratio of the reflection coefficients for vertical and horizontal polarization at oblique incidence is measured. A hand-held instrument for the measurement of for example paint and ceramic coatings on tools and turbine blades has been developed. The thickness range is 4-200 μm and the measurement area is in the order of 2x3 cm^2.

RADAR SENSORS

Sensors that measure the time of flight or the frequency of the echo from an object are called radar sensors. More specifically, the sensor may be a pulse, impulse, FM, or doppler radar, or an interferometer or a combination of those. Radar sensors are used for example for the measurement of surface level in vessels, vibration, movement (burglar alarms, door openers), shape, and for the detection of subsurface reflecting interfaces or objects. An important application is the anti-collision and blind spot radars for cars being developed by several

companies. The first versions are already in use. For example in 1993 all the 2400 Greyhound buses in the U.S.A. were equipped with the Eaton Vorad system (blind spot detector on 10.525 GHz, anti-collision radar on 24.125 GHz).

Autronica AS in Norway and Saab Marine Electronics Ab in Sweden have developed radar sensors for the measurement of surface level in tankers. The sensors can be installed behind a dielectric window to reduce the explosion risk. Because of the target being a single, clearly defined surface, the measurement accuracy is a few millimeters, much better than the theoretical resolution.

Keltronics Ab in Sweden has developed a doppler radar for vibration measurement of large machines in the industry. The control unit contains a software package for analysis of the results for diagnostic purposes.

The applications of impulse radar have been studied in many places [5]. The routine use of impulse radars include detection of pipes and cables in the ground, investigation of the ground for the construction of roads etc., detection of the ground water level, detection of rot in living trees, and investigation of the thickness of the peat layer in marshes. Focusing by aperture synthesis methods have also been studied [6,7]. Focusing provides some attractive features but makes the radar more complicated.

SPECIAL SENSORS

In many cases some specific feature of an object can be utilized in the sensor. For example knots in timber are detected because of their ability to act as dielectric waveguides. The angle of grain in timber is detected from the anisotropy of the permittivity.

At the Belorussian Academy of Sciences in Minsk a hand-held sensor for the inspection of dielectric coatings has been developed [8]. The measurement principle is based on the propagation of surface waves in the coating layer. The sensing element is a dielectric waveguide at a small distance (<1 mm) from the object. For certain frequencies the propagation constant in the dielectric waveguide equals that of the surface waves in the coating, causing strong coupling to occur.

RADIOMETER SENSORS

Microwave radiometers receive the black-body radiation emitted from an object. Because the black-body radiation depends on the temperature, radiometers can be used for the measurement of temperature from a distance through for example smoke or fog, where infrared radiometers will fail. Microwave radiometers can also measure the internal temperature of an object. If several frequencies are used, the temperature as a function of depth can be derived. Medical applications of microwave radiometry have been studied in many institutes. The research at the University of Lille [9] in France has led to commercial aquipment marketed by O.D.A.M. The product is equipment for treatment of cancer by microwave hyperthermia and simultaneous temperature monitoring by microwave radiometry. In Italy several institutes have cooperated in tests with multifrequency radiometry for the detection of subsurface temperature as a function of depth in biological tissues [10].

ACTIVE IMAGING

By measuring both the phase and the amplitude of the field reflected from (holography), or transmitted through (tomography) an object, the three-dimensional distribution of the permittivity can be calculated. These methods are being studied all over the world and the future applications will probably be found in the field of medicine, and in detecting hidden objects in for example security checks. The research in France [11] has led to some commercial products of Satimo. One (the so called microwave camera) is designed for the three-dimensional inspection of biological tissues. Another consists of a linear source and a linear array of sensors (Figure 3). It can measure the permittivity distribution with a resolution of 1 cm of a sheetlike material moving through the sensor.
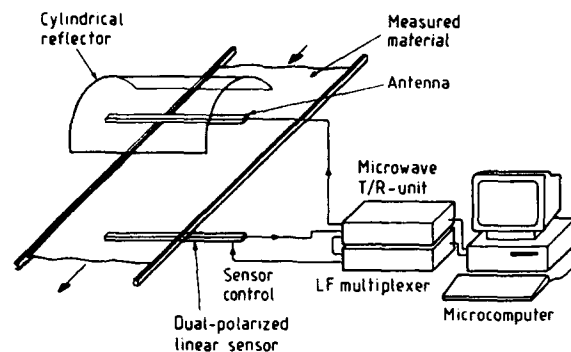
Figure 3. Linear tomographic system made by Satimo for the measurement of sheetlike materials in the industry.

DEVELOPMENT OF MICROWAVE SENSORS IN FINLAND

In Finland industrial microwave sensors have been developed for more than 20 years at the Radio Laboratory of the Helsinki University of Technology and in some companies. Many of the sensors are related to applications in the wood and paper industry.

MEASUREMENT OF SAWED TIMBER

During the late 70´s instruments were developed for the sawmills for the measurement of moisture content, for strength grading, and for automatic edging. The work was performed as a cooperation between Innotec Inc. and the Radio Laboratory of Helsinki University of Technology.

The moisture gauge contains several sensors [12]. The microwave part is a free-space transmission sensor, which measures the microwave attenuation through the timber at 10 GHz. The density is measured with a gamma-ray attenuation sensor and the temperature with an infrared radiometer. The moisture is derived from combining the results.

The strength grading of timber is based on the measurement of the dry-density, knottiness, and slope of grain [13]. From these parameters the strength is calculated using an empirical model. The density is measured with a gamma-ray attenuation sensor. To get the dry-density the moisture has to be subtracted. It is measured with a microwave free-space transmission

sensor and an infrared radiometer. The slope of grain is measured by detecting the turning of the polarization of a linearly polarized microwave signal travelling through the timber. The sensor measuring the size and location of the knots [13] is a microwave bridge based on free-space transmission through the timber in both channels.

The equipment for the automatic edging of boards (a slice from a log) contains some optical sensors and a microwave sensor for detection of knots. This knot detector illuminates the board from above with a plane wave. Underneath is an array of cylindrical waveguide detectors. The knots act as dielectric waveguides matching the plane wave to the mode TM_{01}, which is detected.

MEASUREMENT OF VIBRATION

A microwave radar for the measurement of vibration in power transmission lines was developed at the Radio Laboratory. The vibrations are caused by the wind and will cause fatigue damages if not damped. The radar is a phase-locked Doppler radar, i.e. an interferometer working at 16 GHz. Because the vibrations depend on the wind, the measurement must be made over a period of time during different weather conditions. Traditional contacting sensors are not suitable for the purpose because of the high voltage.

PROBES FOR MATERIAL MEASUREMENT

At the Radio Laboratory probes have been developed, which can be pushed into soft materials. The "snow fork" (Figure 4a) for the measurement of the dry density and the liquid water content of wet snow is a quarter-wavelength resonator made of two-conductor line [14]. The resonant frequency and the quality factor are measured, from which the density and the moisture are calculated. The "peat probe" (Figure 4b) for the measurement of the energy content of peat in marshes is a slot resonator made of a steel tube [15]. The resonant frequency correlates directly with the energy content. Both the snow fork and the peat probe are manufactured by Toikka Engineering.
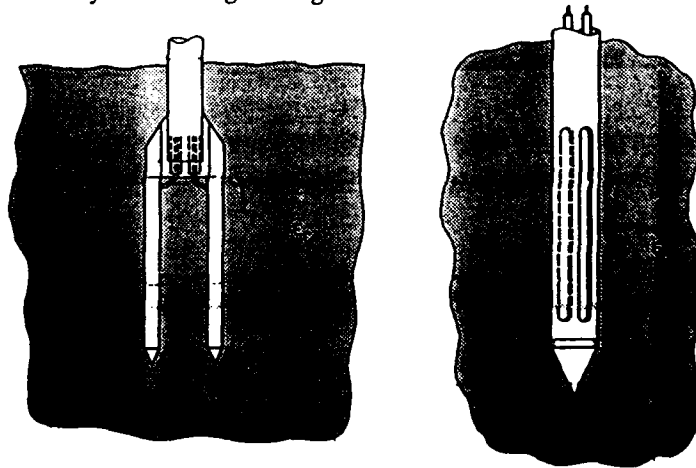
Fig. 4. The snow fork (left) and the peat probe. The peat probe is filled with epoxy resin. Both can be used with other materials like liquids, powders, and granular materials.

MEASUREMENT OF SPEED AND TRAFFIC

At Ylinen Electronics a range of Doppler sensors have been developed. One is a speedometer sensing the speed of an icebraker relative to the ice. Another is a portable traffic counter, which can easily be moved from one location to another. It is attached to a pole by the side of the road. The counter can be used on small gravel roads also, where the use of inductive loops is impossible. The operating frequency of both instruments is 34 GHz.

MICROWAVE CONSISTENCY ANALYZER

A sensor for the measurement of the consistency of pulp has been developed at Valmet Automation. There are two versions of the sensor, both of which are free-space transmission delay sensors mounted to a steel pipe (Figure 5). The operating frequency is 2-3 GHz. The sensor can be used with high consistencies and is not affected by species of wood, freeness, fiber length, flow speed, or kind of pulp. The consistency measurement range of the instrument is 1-15 % with an accuracy of 0.1 %-units.
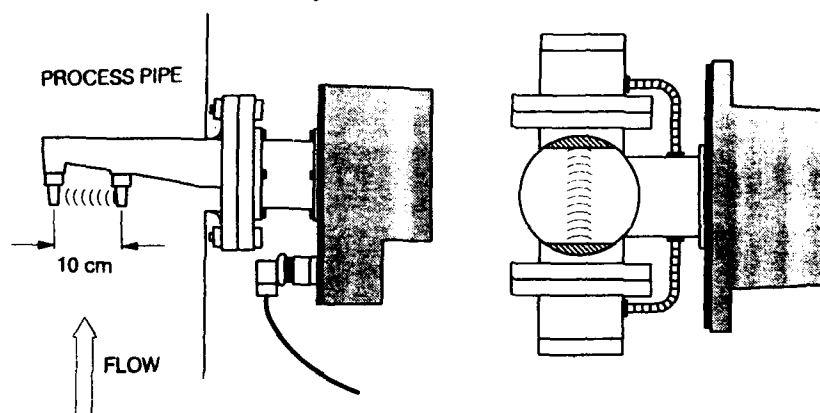
Figure 5. The microwave consistency analyzer is made in two versions, one for thick pipes (left) and one for thin pipes.

APPLICATIONS OF IMPULSE RADAR

In Finland the first application of the impulse radar was the measurement of the thickness of the peat layer in marshes for inventory purposes [15]. The frequency of the radar was 100 MHz. Due to the homogeneity and the high permittivity (70-80) of the peat, the thickness resolution was as good as 10-20 cm. The next application was the detection of rot in living trees with a 900 MHz radar. In these applications, U.S.-made radars were used. There was, however, a need for a higher-frequency radar with better distance resolution in dry objects (low permittivity). Therefore a 2500 MHz impulse radar was developed at the Radio Laboratory. The equipment is compact, portable, battery operated and controlled by an industry standard computer [16]. The HF unit contains a step-recovery diode transmitter and two quasi-TEM antennas. In the application tests of the radar prototype, embedded structures of buildings (beams, reinforcing bars, air channels) and rot in living trees and wooden telephone and power line poles have been detected. The radar came onto the market in 1993 (Toikka Engineering).

RESONATOR SENSOR FOR THE MEASUREMENT OF HUMIDITY OF AIR IN DRYERS

Water vapour increases the permittivity of air. This small (typically less than 0.3 %) change can be measured by detecting the decrease of the resonant frequency of a microwave resonator. This kind of resonator sensor was developed at the Radio Laboratory for the measurement of the humidity of air in a veneer dryer in a plywood factory [17]. The resonant frequency of the sensor is about 10 GHz. It is a TE_{011}-mode cylindrical resonator, whose end structures can be quite open letting air flow freely. The TE_{011}-mode is advantageous also in that sense that the electric fields are zero on all surfaces of the resonator and contamination of dirt has therefore only a minor effect. The temperature in the dryer is high (up to 200 °C) and there is dirt and resins in the air preventing the use of e.g. capacitive humidity sensors.

STRIPLINE RESONATOR SENSORS AND THEIR APPLICATIONS

Stripline resonator sensors have been developed at the Radio Laboratory from the beginning of the 80's. These sensors are suitable for the measurement of the properties of planar dielectric objects. The material under test is used as part of the dielectric of the stripline. which can have one or two center conductors (and two ground planes). One advantage compared to microstrip sensors is the much higher quality factor (up to 3000). A half-wavelength VHF stripline sensor was developed at the Radio Laboratory for the measurement of the mass per unit area of a layer of wooden particles (thickness 40-120 mm) in a particle board factory [18]. Fast mapping of the properties of the 1.5-3 m wide layer is realized with a linear array of 10-20 adjacent sensors. To minimize the effect of mutual coupling between sensors, pin-diode switches are used on the strips to switch the resonators on and off. One common measurement equipment is used for the array, and thus the system is rather inexpensive.
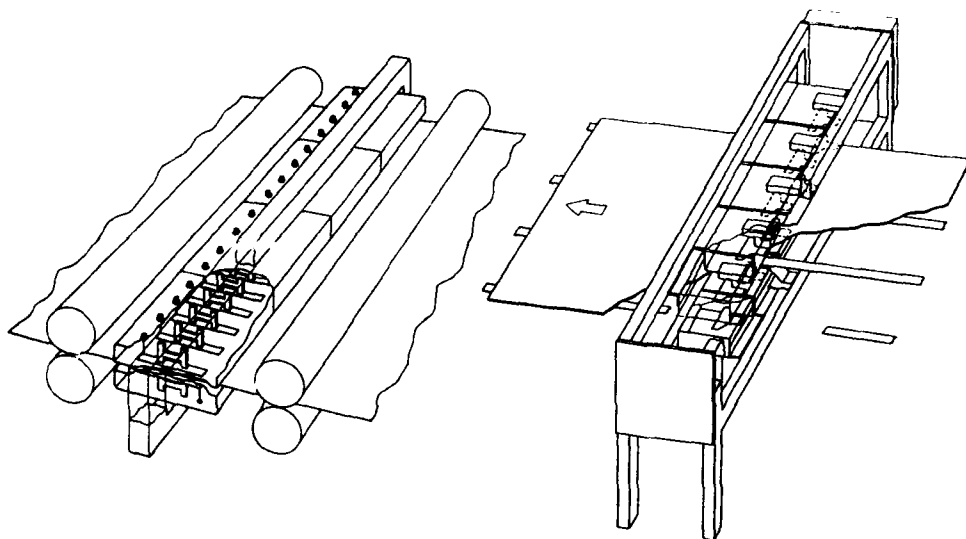
Figure 6. The stripline resonator array for the measurement of the water content of wet paper web (left) and for the measurement of veneer sheets.

For thinner dielectric layers, more sensitivity is obtained by using the stripline resonator sensor with two center conductors and two ground planes. The structure can support two resonant modes, even and odd, whose resonant frequencies differ about 10 %. The electric fields are orthogonal in the plane of the sample (the symmetry plane) and thus have different responses to the permittivity of the sample. This makes error compensaton possible. This idea was utilized in a sensor for the moisture measurement of wet paper web in a paper machine [19]. By using an array of more than 100 half-wavelength 400 MHz sensors spaced 0.1 m apart, fast mapping of over 10 m wide webs is possible (Figure 6). The measurement electronics can measure 100-200 sensors per second, which makes the measurement of the moisture profile an almost real-time measurement. Another application of the stripline resonator sensors with two center conductors is the two-parameter measurement of both moisture and density of veneer sheets in a plywood factory [20]. In this measurement, quarter-wavelength resonators are used (Figure 6). These support only the even mode, but now both the perturbation of the resonant frequency and the quality factor are measured to determine the real and imaginary part of the permittivity. From these the moisture and the density are calculated.

At the present the applications of scattering in inhomogeneous media, the time-dependence of phenomena related to bound water, and the applications of stripline sensors and radars are studied at the Radio Laboratory.

REFERENCES

[1]   Nyfors, E., P. Vainikainen, "Industrial Microwave Sensors", Norwood, MA: Artech House, 1989, 351 p.

[2]   Klein, A., W. Pesy, "Experiences with the microwave moisture meter 'Micro Moist'", Mineral Processing, Vol. 30, No. 9, 1989, pp. 549-557.

[3]   Kalinski, J., "On-line coal-dust moisture content monitoring by means of microwave method and instrumentation", Proc. 18th European Microwave Conf., Budapest, Sept. 1990, pp. 1673-1678.

[4]   Konev, V., N. Lyubetsky, S. Tikhanovich, "Non-destructive testing of materials by microvave ellipsometry methods", Proc. 12th Int. Conf. on Non-Destructive Testing, Amsterdam, 1989, pp. 1630-1632.

[5]   Proc. 4th Int. Conf. on Ground Penetrating Radar, Rovaniemi, Finland, June 1992.

[6]   Daniels, D., D. Gunton, H. Scott, "Introduction to subsurface radar", IEE Proc. F (Special issue on subsurface radar), Vol 135, No. 4, Aug. 1988, pp. 278-320.

[7]   Junkin, G., A. Anderson, "A new system for holographic imaging of buried services", Proc. 16th European Microwave Conf., Sept. 1986, Dublin, pp. 720-725.

[8]   Konev, V., V. Mikhnev, "Inspection of the parameters of sheet dielectrics by analysis of the frequency properties of a dielectric waveguide sensor", Sov. J. Nondestr. Test. (US), Vol. 22, No. 6, June 1986, pp. 367-372.

[9] Bocquet, B., J. van de Velde, A. Mamouni, Y. Leroy, G. Giaux, J. Delannoy, D. Delvalee, "Microwave radiometric imaging at 3 GHz for the exploration of breast tumors", IEEE Trans. Microwave Theory Tech., Vol. 38, No. 6, June 1990, pp. 791-793.

[10] Bardati, F., G. Calamai, M. Mongiardo, B. Paolone, D. Solimini, P. Tognolatti, "Multispectral microwave radiometric system for biological temperature retrieval: Experimental tests", Proc. 17th European Microwave Conf., Rome, Sept. 1987, pp. 386-391.

[11] Pichot, C., L. Jofre, G. Peronnet, J.-C. Bolomey, "Active microwave imaging of inhomogeneous bodies", IEEE Trans. Ant. Prop., Vol. AP-33, No. 4, April 1985, pp. 416-425.

[12] Tiuri, M., K. Jokela, S. Heikkilä, "Microwave instrument for accurate moisture and density measurement of timber", J. Microwave Power, Vol. 15, No. 4, 1980, pp. 251-254.

[13] Heikkilä, S., P. Jakkula, M. Tiuri, "Microwave methods for strength grading of timber and for automatic edging of boards", Proc. 12th European Microwave Conf.., Helsinki, 1982, pp. 599-603.

[14] Sihvola, A., M. Tiuri, "Snow fork for field determination of the density and wetness profiles of a snow pack", IEEE Trans. Geoscience and Remote Sensing, Vol. GE-24, No. 5, Sept. 1986, pp. 717-721.

[15] Tiuri, M., M. Toikka, I. Marttila, K. Tolonen, "The use of radio wave probe and subsurface radar in peat resource inventory", Proc. Symp. IPS Commission I, Aberdeen, 1983, pp. 131-143.

[16] Vainikainen, P., M. Tiuri, V. Kontra, M. Saarikoski, E. Nyfors, R. Salminen, "High-resolution portable impulse radar", Proc. 19th European Microwave Conf., London, Sept. 1989, pp. 1091-1095.

[17] Toropainen, A., P. Vainikainen, E. Nyfors, "Microwave humidity sensor for difficult environmental conditions", Proc. 17th European Microwave Conf., Rome, Sept. 1987, pp. 887-891.

[18] Vainikainen, P., E. Nyfors, "Sensor for measuring the mass per unit area of a dielectric layer: Results of using an array of sensors in a particle board factory", Proc. 15th European Microwave Conf., Paris, Sept. 1985, pp. 901-905.

[19] Fischer, M., P. Vainikainen, E. Nyfors, "Dual-mode stripline resonator array for fast error compensated moisture mapping of paper web", IEEE MTT-S Int. Microwave Symp. Digest, Dallas, May 1990, pp. 1133-1136.

[20] Vainikainen, P., E. Nyfors, M. Fischer, "Radiowave sensor for measuring the properties of sheetlike dielectric material: Application to veneer moisture content and mass per unit area measurement", IEEE Trans. Instrumentation and Measurement, Vol. IM-36, No. 4, Dec. 1987, pp. 1036-1039.

Theory and Applications of Polarimetry in Radar

Robert T. Hill
Consultant
Bowie, Maryland, USA

Abstract: This paper, invited for the conference MIKON-94 in Ksiaz, Poland, provides a survey of recent developments in the uses of polarimetry in radar, a subject the author feels represents one of the exciting "new directions" in radar development. The survey provides a brief review of theory and then covers three applications (non-exclusive) that have been the context of recently reported work: polarimetry as a CFAR-detection mechanism, polarimetry as a means of target classification, and polarimetry used with other techniques (e.g., SAR and ISAR) to improve imagery. The paper draws entirely upon recent open literature, therefore providing a useful reference base for those intending further study or wishing to contact the researchers.

INTRODUCTION

In the later years of my career in radar development, and since my retirement, I have lectured a great deal about current developments and "new directions" in radar. Today, these "directions" are no longer "new" perhaps, but they still represent the distinct change in radar that took place in the 1970's (approximately). This change, in my view, involved abandoning the "radar performance factor" (that ratio of pulse power generated in the transmitter to the minimum discernable signal of the receiver) as the dominant measure of relative value among radars (it having reached 200 dB routinely), and adopting instead several somewhat more subtle qualities that do not involve power and sensitivity so dominantly. These qualities, the vigorous pursuit of which was enabled by great advances being made in computers and computer-like apparatus at the time and in the advent of inertialess beam steering in radar, I called [1]:

* Information Content of Signals
* Efficiency through Flexible Operations
* Completing the Estimation, the Report.

I further spoke to these directions in the Chinese conference in 1986 [2] and again in China in an unpublished presentation in 1991. These qualities relate, I'm sure, to increased emphasis in radar upon "system engineering", demanding that radars be designed as a part of the system they serve, reminding us in turn that radars are but sensors which form as their output estimations, always with error, of the true state of nature at their input.

In this lecture, I address developments in polarimetry that I find particularly fascinating and which I consider to be primarily a part of the first of these directions listed. Of course, the other two qualities permit the sensor to devote necessary dwell time and to construct its

output in a user-significant way; in general, such advances in radar do not lie exclusively in any one of these directions.

IMPORTANCE OF POLARIMETRY

Radar sensing differs from human experience in using electromagnetic waves (optical vision) in two specific ways - humans are directly sensitive to neither coherence nor polarization. Our vision evolved in sunlight, which is neither polarized nor narrow-band enough to afford other than just occasional experience with its wavelike properties (e.g., observed prismatic effects and interference - rainbows and Newton rings, say). Yet in radio physics, these properties are immediately obvious and were so, in a sense, from sometime between Faraday's experiences (about 1831) and the vector field accounting of Maxwell (1864). Constraining an alternating current (with requisite acceleration of charge) to the path of a conductor (e.g., a wire as an antenna) defines the orientation of the electric-field constituent of the emanating electromagnetic wave, and it is this field orientation that we call the "polarization" of the wave.

The importance of this property of radio (hence radar) waves lies in the fact that the scattering of the waves from objects of various substances, shapes and orientation is a function of the wave's polarization. We have long recognized this dependence in such familiar areas as the forward scattering from Earth's surface, resulting in multipath interference, and in some measures of clutter reduction. Today, importance lies in the more exhaustive treatment of this dependence in the backscattering experienced and the hypothesizing of what target conditions account for it. One exciting result is that it is quite possible to discriminate among targets by tests on the space of polarimetric data without fundamental dependence upon the relative signal strengths, a massive departure from historic "thresholding" in radar detection.

OVERVIEW OF THEORY

Figure 1 illustrates the point made earlier, that we in radar experience the polarization of an electromagnetic wave most fundamentally when we visualize the source current constrained to a specific orientation, and the orientation of the electric field is aligned; it is this "E-field" orientation that we cite as the whole wave's "polarization". (We let this electric-current-source case suffice for showing "existence" without examination of any magnetic dual as a source.)

One must appreciate that the general polarization of an electromagnetic wave (an "EM wave") is elliptical; that is, any consistent E-field orientation is a form of an ellipse (certainly including the cases of "linear" and "circular" polarization). In my teaching I use the illustration of Figure 2 to show this; it shows crossed dipoles (constrained current paths - one horizontal, one vertical) and the phasing between them that would result in the "circular" polarization shown (i.e., to an observer at the field point shown the E-field is apparently rotating with unchanging strength in the "right hand" sense illustrated). This is a sufficient source model since by adjusting the relative amplitudes of the two currents (including letting one vanish) and

the phase between them, one can establish any linear, elliptical and circular (i.e., any elliptical) polarization, the complete set. The "polarization ellipse", then, shown in Figure 3, gives a complete account of all possible cases and identifies two parameters, the ellipticity angle tau and the polarization angle phi, which, when the sequencing "sense" is noted (right or left hand, by convention), are sufficient to describe the polarization of an EM wave.

A vitally important "mapping" of these parameters is shown in Figure 4, the "Poincare sphere", points on the surface of which represent the set of all polarizations; the sign of tau indicates the rotational sense, determining in which hemisphere is the plot. We note,too, the mapping results in orthogonal polarizations being represented by antipodal points (note, for example, the horizontal and vertical polarizations, or the left- and right-handed circular "poles"). Polarization unchanging over at least a brief time period is necessary in these definitions, since even in "unpolarized" waves (sunlight) there is at any instant some orientation of the E field; a brief averaging process would show an interior point in the Poincare sphere, a volume reserved, then, for "partial" polarizations, the center for "unpolarized" waves. Regions outside the unit-radius sphere allow us to account for error in estimating (sensing) the true polarization of a wave.

So, a single point on the Poincare sphere tells us the polarization of a wave, but nothing about a target. We now briefly examine scattering and elementary "polarimetry" at the same time. In Figure 5 we see, in a matrix noted [S], an account of the polarization of the waves returned from a radar target (this now in the backscattering context pertinent to monostatic radar) as a function of the incident wave's polarization. For example, had we transmitted vertical polarization and measured both the vertical ("co-polar") and horizontal ("crosspolar") signal components received, and noted them as the subscripts indicate, and then transmitted horizontal polarization and done the same, we would have filled this matrix with four terms, each complex (amplitude and phase noted). Since the size and range to the radar (in wavelengths) do not bear upon a target's "polarization character", this "scattering matrix" contains three independent terms (six variables) after a normalization of signal strength and disregard of path phase length.

Measurements made in any pair of orthogonally polarized receiving "channels", first, from a signal transmitted on one polarization and then from a second transmission orthogonal to the first are adequate to "fill in" the entire "scattering matrix" [S], but often the apparatus used gives us on receive the cross- and copolar relationships to the transmitted signal assumed in the notation of Figure 5.

The scattering matrix [S] completely defines the character of the target at the time of the scattering. It is not too difficult to relate with familiar trigonometric identities the measurements recorded in [S] and estimates of the polarization vector (a point near the Poincare surface) for each of the return waves from the separate transmissions; such is reviewed in [3]. With that mapping, we see that two plots on the Poincare sphere, one for each transmission are required to represent the the scattering matrix associated with a target.

CURRENT RESEARCHERS

I credit those working in the radar specialty area of "remote sensing" for the impressive work reported in the open literature about radar polarimetry. A splendid review presented by Dr. F. Ulaby of the University of Michigan in the 1989 radar (and other imaging sensor) conference in Kyoto, Japan, was represented in the record by only a brief abstract - but a helpful survey paper by Wolfgang Boerner (a frequently referenced contributor) and Y. Yamaguchi was also on that program [4]. The orderly treatment of polarimetry given by Dr. Dino Giuli in 1986 is also cited frequently [5]. In this survey I draw on work reported by researchers in Germany (Wanielik and Stock), France (Pottier and Saillard; Durand et al), Denmark (Krogager) and the U.S., cited in the paragraphs to follow. I am particularly indebted to two U.S. researchers, L.M. Novak at the Lincoln Laboratory of MIT and Fred Garber at Wright State University in Ohio, who have chosen to sustain a personal communication with me as their efforts continue. I find it convenient to represent the work of these many researchers in three application areas: CFAR detection, target classification, and image enhancement.

CFAR DETECTION

Detection in radar involves the well-known statistical process of hypothesis testing, in which "target" and "no target" hypotheses are tested on the data space of an "observable" of the receiver's output; a statistically optimum test (not proven here) can be based on the relative likelihood of the hypotheses given the data observed and certain knowledge of the statistics involved. Generally, the problem of detection of a reasonably well-known signal in the presence of radio noise reduces to that of a "threshold" test on the amplitude of the receiver's video output. Such a test is illustrated in Figure 6, very familiar to all students of radar and radio detection. The frequency with which one produces false alarms (declaring a target to exist when none does) is determined by the absolute threshold setting and the shape of the probability density function (pdf) actually indicative of the "null" state (i.e., the no-target or clutter-only state). If the radar environment contains great changes or other uncertainty in the very statistics of the clutter, then a way of setting the threshold relative to that clutter signal is necessary to make constant an acceptable probability of false alarm, P_{FA}. Receivers in which the statistics of the background are estimated and the threshold then set accordingly are called Constant False Alarm Rate (or "CFAR") receivers . . and the frequently used observable is the video output amplitude illustrated here.

Polarimetry provides another observable entirely for such thresholding. A receiver, it is reasoned, could "learn" the statistics of the polarization of the signals in a locality of resolution cells and upon encountering a cell in which the polarization of the return lies outside the statistically bounded region of the "background" (no-target state), that is, when a "polarization threshold" is crossed, a target can be declared - the cell apparently does not "belong" to the background. Ground clutter rejection by polarimetric means was described by Giuli et al in 1989 [6] and polarimetric CFAR was further reported by

Wanielik and Stock in 1990 [7] from which I have taken Figure 7 to illustrate the idea. This shows the very terms of the scattering matrix [S] being examined for departure from the backgound; Figure 8 shows experimental data for rain, land and an aircraft (a Dassault Falcon 323). Here the data are represented by plots on the Poincare sphere ("unfolded" in these Molleweide projections of the sphere). The left column of plots is for the rain, the center column the land and the right the aircraft. The top row is the experience when horizontal polarization was transmitted, the middle row for vertical and the bottom row for left circular. The top two rows are sufficient in the sense that [S] is fully represented, the third row corroborates. The distinctions among the target types is fairly obvious. Note that this particular kind of land is acting somewhat like a "polarization filter" - consistent polarization is returned for any incident polarization without regard to whether the return was particularly strong or not, and also that the aircraft return is consistent and, more to the point, well removed in this polarization space from either of these two backgrounds.

CFAR detection was also the subject of the report by Pottier and Saillard [3] from which I've taken Figure 9 here to represent the thresholding idea more clearly and Figure 10, a collage of figures used in the reference which relates an interesting anechoic-chamber experiment. In this experiment, the circuits indicating the components of the polarization vector were monitored, one such recorded here, while a reflecting "net" was illuminated and being shaken slightly to afford some variation (still statistically stationary). Then, a small metallic sphere was dropped in front of the net and the disturbance in polarization but NOT in power is easily seen by the two charts shown. The authors remark that "In the polarimetric CFAR detection, the originality resides on the fact that at no time (does) the process take into account the received power, but uses only the discrimination between different polarization states that occur during the dectection."

A further refinement to P-CFARing (with possible application to target recognition as well, it seems to me) was presented in 1993 by J.C. Durand of Thomson-CSF at the U.S. National Radar Conference [8]; the "migration" of the polarization vector in short periods of time (as little as 68 ms recorded) from wooded areas subjected to rather strong winds is illustrated in Figure 11. Here we see "filaments" on the Poincare sphere representing five cells of such terrain, a fixed radio tower among them. I believe that such short-term temporal behavior could be a part of future "signature analysis" in target classification.

TARGET CLASSIFICATION

Of course, polarimetric CFAR detection involves "target classification" based of polarimetry, the classification being targets not of interest (clutter) and those of interest (aircraft). It is hard to imagine that all the world's aircraft might occupy such well contained and unique spaces on the Poincare sphere that inter-aircraft classification could be reliably achieved. However, the work reported by Chamberlain, Walton and Garber [9] [and by the several references cited therein] introduces polarimetry in which returns from the principal scatterers composing a target in the down-range dimension are resolved.

The local polarization vector, then, down the length of a target forms a "transient polarization response" (TPR), illustrated in Figure 12. Figure 13 (adapted from [9]) shows that a number of the points, ordered from front to back (here 1-4) and plotted on a "polarization chart" (the convenient projection of Poincare sphere plots onto the equatorial plane, with rotational sense requiring a coding as shown) makes an interesting "signature" of the aircraft, possibly sufficiently separated from others (the two shown in Figure 13 give the notion) to contribute to identitfication (i.e., more definitive than just classification).

An interesting observation here is that radar methods of target recognition fall into two categories: (1) the popular developments in 2D "imaging" using (usually) inverse synthetic aperture radar (ISAR) techniques so that the basis for recognition is near-visual perception by the human, and (2) the use of far more "abstract" data spaces, such as we illustrate here. The irony is that while the visual imaging makes a "reference library" easy to acquire, it may be asking the radar to do something (form an image) for which it is not particularly well suited, whereas taking better advantage of what radio can better do (sense polarization, for example) results in data spaces less familiar to the human experience and for which a sufficient reference base is far harder to acquire. I believe that system design in the future should try to accommodate the more abstract data spaces related to what the sensor does best and not depend only on imaging in the literal (visual) sense.

There are other "classifaction" efforts using polarimetry in which the objective is to classify the geometrical type of scatterer involved. Cameron and Leung described in 1990 [10] a "classification of scattering matrices" forcing an estimation into one of some eleven geometries (right helix, left helix, trihedrals, diplanes, cylinders, etc.). Such estimation could be integrated with other processes, it seems to me, such as the tagging of the dominant scatterers isolated in the downrange TPR processing, as it has been in the categorizing of the scatterers isolated in SAR and ISAR imaging, our next subject.

IMAGE ENHANCEMENT

In addition to the scatterer classification of Cameron and Leung just mentioned, an interesting decomposition of the scattering matrix measured by polarimetric radar has been described by Ernst Krogager of Denmark [11]. Whereas others (Huynen [12] among them, Krogager points out) have formulated the decomposition of a form of [S] into a "symmetrical" target matrix and a (roughly) "noise" target, Krogager presents a decomposition into only three scatterer types: a sphere, a diplane and a right or left wound helix. This process is then applied to polarimetric synthetic aperture radar (SAR) imaging.

The author first examines polarimateric measurements simulated for a target composed of various geometric forms to illustrate what would be lost in single-polarimetric SAR imaging and how resolution of scatterers is increased by separate "imaging" in seven treatments: HH, HV, VV, k_{sphere}, $k_{diplane}$, k_{helix} and "maxpol", where the k terms are from the subject decomposition and maxpol refers to a Huynen parameter giving a true intensity of the scatterers. The author further illustrates

application to fully polarimetric SAR data made available by German and U.S. agencies associated with measurements made at Oberpfaffenhofen, Germany. He examines this land scene, with vegetated areas and man-made structures, including rail facilities, using each of the "images" formed, and shows what added inference the "k" images permit.

Among others investigating use of fully parametric SAR measurements is L. M. Novak (M.I.T. Lincoln Laboratory). In [13], several detection "tests" are compared with attention to effectiveness and need for a priori data. In [14] he describes, first, the Lincoln Laboratory 33 GHz SAR (with 1 ft. x 1 ft resolution), then the Polarization Whitening Filter (PWF) methods used to minimize clutter deviation σ_c - clutter speckle. The paper produces some remarkable images which I will not trust to reproduction here but will use in the lecture; here I will reproduce only the histogram comparison of single channel imagery with that employing PWF, Figure 14, involving a powerline-tower scene. One can see the reduction of the background spread σ_c and can infer the increased contrast (detectability) of the towers relative to the background, clearly evident in the images presented. Both [13] and [14] include the formulation of PWF weights applied.

In a paper last year [15] Dr. Novak applies PWF to scenes in which militarily-significant targets are present and compares results of using polarization matched filtering (PMF, in which complex weights can be applied to HH-HV-VV data based on target hypotheses to provide a maximum average target-to-clutter ratio T/C) with and without PWF. While some loss in the PMF effect is incurred when PWF is employed jointly (this loss is tabulated for various cases), the paper concludes that for the cases at hand the HH-VV (+45° tr., -45° rec.) was the most effective polarization for the PMF technique and that PWF significantly increased the (T/C)/σ_c for further improvement in the CFAR detection. In [14] is further discussion of adaptive PWF weight determination and some attention to homogeneity assumptions.

SUMMARY

The last few paragraphs have shown indeed that the three "subjects" (CFAR detection, target recognition, and image enhancement) here reviewed with regard to the contribution made by polarimetry are not at all mutually exclusive; it was a convenience here to present the work of various contributors in that sequence. Ultimately, I suppose, all sensing has as its objective the least error in estimating the true state at the sensor input; all three subjects of (1) detecting the mere presence of an aircraft, perhaps of very little radar cross section, (2) classifying (even identifying) such a target, and (3) correctly interpreting an entire scene as in "remote sensing" for many purposes could be contained in the term "detection", used so much in the radar literature. "Estimation" is the term I prefer, reminding us of the stochastic nature of sensing.

We have explored here estimation by the use of microwaves as the medium and radar as the sensor. Work in radar polarimetry represents to me an appropriate exploitation of the properties of that medium and that sensor, and one that could easily have been ignored (possibly was being

ignored) because it is not common to the human experience. Techniques
now exist to enable this exploitation. The further implication of
sensor "fusion" is, I believe, obvious, wherein several sensors of
differing properties, each "exploited" fully, would be used to reduce
still further uncertainty (error) in the joint estimation.

It may be that, one day, synthetic visual "scenes" may be produced
in three spatial dimensions from such fused information, but I do NOT
hold to the necessity of that. In fact, I encourage all sensor develop-
ers, and the "user" community as well, to become comfortable with far
more abstract spaces upon which extraordinarily accurate estimations
might best be presented. Certainly the inner workings of such proces-
sing will demand this accommodation.

One of the reasons I teach often about radar polarimetry is that I
so admire those doing the work. To them, those cited in this summary
and those not, I say my thanks now, and lend my further encouragement.

<p align="center">*   *   *   *   *</p>

REFERENCES

[1] Hill, R. T., "Evolution of Radar: a U.S. Viewpoint", Colloque
     International sur le Radar, Paris; SEE, May 1984

[2] Hill, R. T., "Radar Tenacity, Adaptivity and Capacity and the
     Importance of Feedback Control", Chinese International Conference
     on Radar, Nanjing; CIE, November 1986

[3] Pottier, E. and J. Saillard, "Optimal Polarimetric Detection of
     Radar Target in a Slowly Fluctuating Environment of Clutter",
     International Radar Conference, Washington, D.C.; IEEE, May 1990

[4] Boerner, W. and Y. Yamaguchi, "A State of the Art Review in Radar
     Polarimetry and its Applications in Remote Sensing", International
     Symposium on Noise and Clutter Rejection, Kyoto; IEICE, November
     1989

[5] Giuli, D., "Polarization Diversity in Radars", Proceedings of IEEE,
     v. 74, no. 2, February 1986

[6] Guili, D., M. Gheradelli and A. Freni, "Adaptive Polarization for
     Rejection of Ground Clutter", Colloque International sur le Radar,
     Paris; SEE, April 1989

[7] Wanielik, G. and D.J.R. Stock, "Measured Scattering-Matrix-Data and
     a Polarimetric CFAR-Detector Which Works on This Data", same
     record as [3]

[8] Dujardin, E., J.C. Durand and B. Carrara, "Polarimetric Features of
     Ground Clutter", National Radar Conference, Boston; IEEE, March
     1993

[9] Chamberlain, N., E. Walton, F. Garber, "Radar Target Identifi-
cation of Aircraft Using Polarization-Diverse Features", Trans-
actions on Aerospace and Electronis Systems, IEEE, v. 27, no. 1,
January 1991

[10] Cameron, W. and L. Leung, "Feature Motivated Polarization
Scattering Matrix Decomposition", same   record as [3]

[11] Krogager, E., "Decomposition of the Sinclair Matrix into
Fundamental Components with Application to High Resolution Radar
Target Imaging", [paper provided by the author; affiliation:
Danish Defence Research Establishment, Copenhagen]

[12] Huynen, J., "Phenomenological Theory of Radar Targets", Doctoral
Thesis, Technical University, Delft, The Netherlands, 1970

[13] Chaney, R., M. Burl and L. Novak, "On the Performance of
Polarimetric Target Detection Algorithms", same record as [3]

[14] Novak, L. and C. Netishen, "Polarimetric Synthetic Aperture Radar
Imaging", International Journal of Imaging Systems and Technology,
v. 4, p. 306; John Wiley and Sons, Inc., 1992

[15] Novak, L. and S. Hesse, "Optimal Polarizations for Radar Detection
and Recognition of Targets in Clutter", same record as [8]
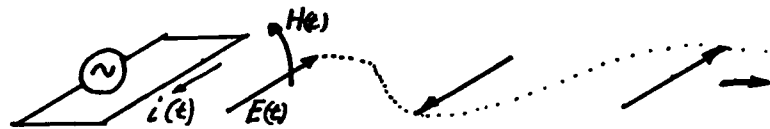
Figure 1   Source Current and Electric Field Alignment - Polarization
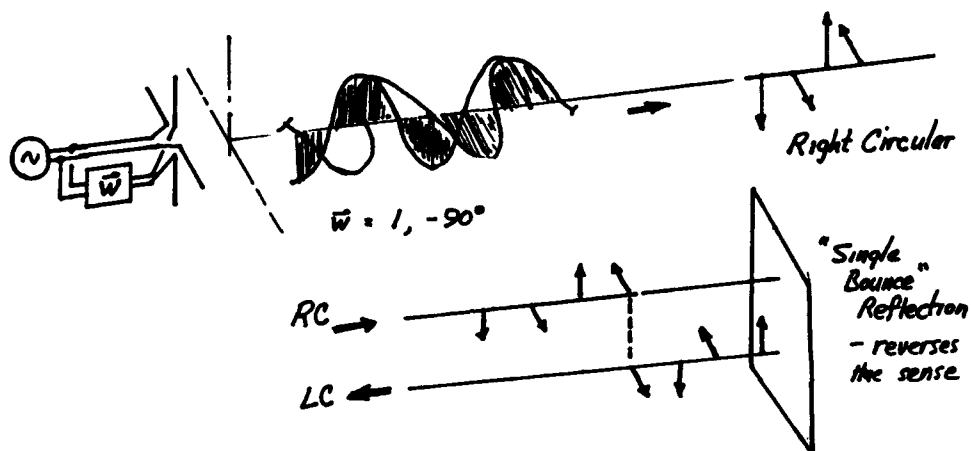
Figure 2   Polarization Model. Showing Reflection
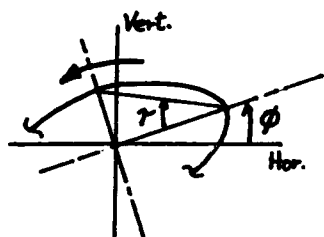
Figure 3   Polarization Ellipse
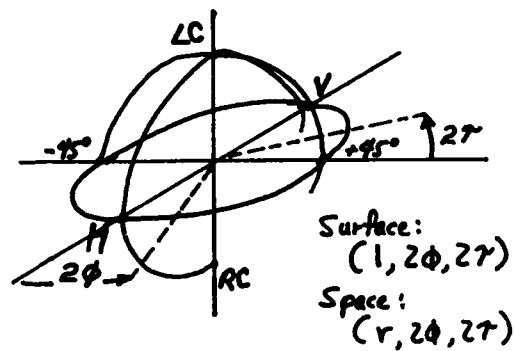
Figure 4   The Poincare Sphere

$$[S] = \begin{vmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix}$$

Recall: S_{ij} are complex

i,j represent orthogonal
polarizations, transmit
and receive respectively

Figure 5   The Scattering Matrix
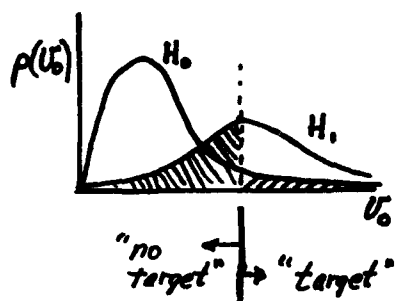
H_O : Noise only
H_1 : Signal and Noise

==> probability of miss
==> probability of false alarm

Figure 6   Detection Test, on Space of
Amplitude of Detector Output

surrounding signal vectors ⟹ $\underline{j}_c$

$\underline{s}(N+1)$ signal vector under test
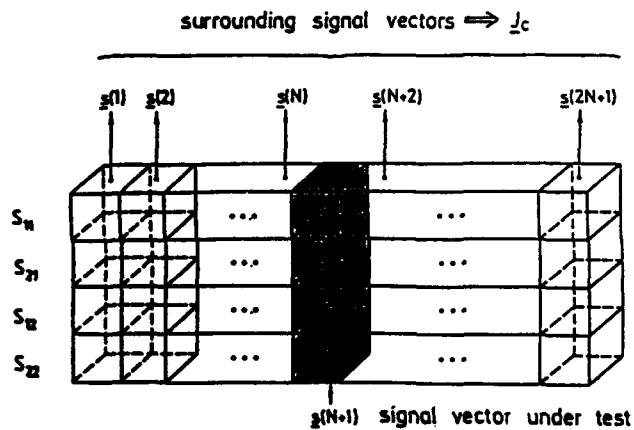
Figure 7   Polarimetric CFAR (from [7])
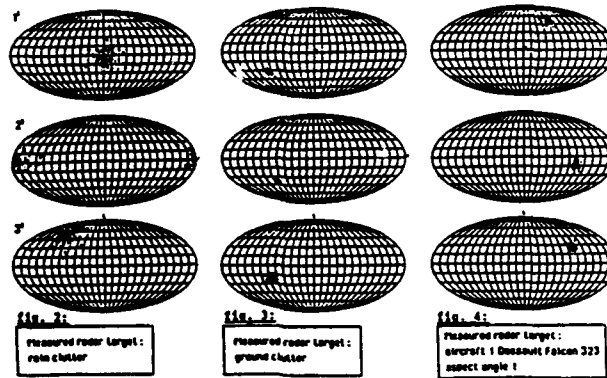
Figure 8   An Experiment: Polarimetric Separation
of Rain, Land, an Aircraft (from [7])
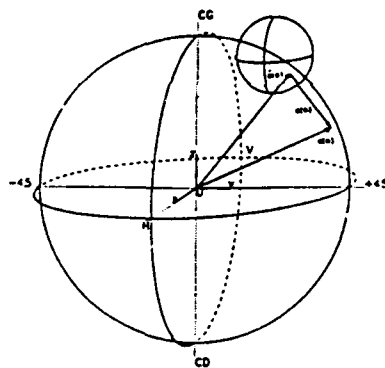
Figure 7: "DECISION SPHERE": CASE OF DETECTION.
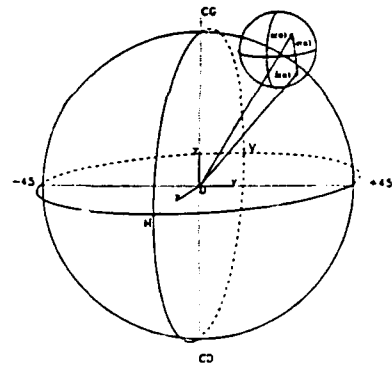
Figure 8: "DECISION SPHERE": CASE OF NO-DETECTION.

Figure 9   Polarimetric Thresholding (from [3])
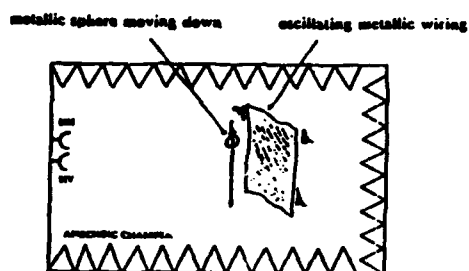
Figure 9: EXPERIMENT CONFIGURATION.
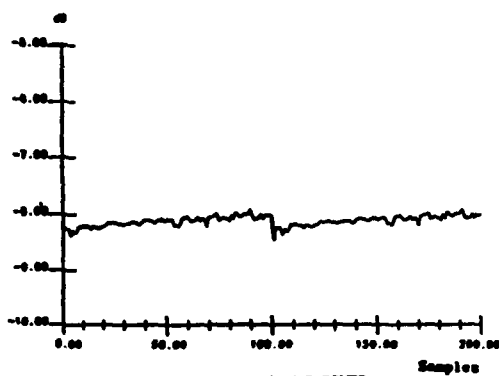
Figure 10: RECEIVED POWER.
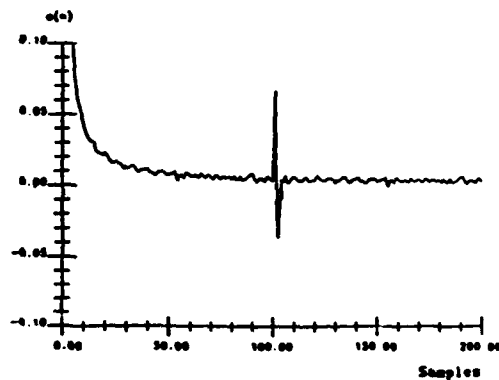
Figure 11: PREDICTION ERROR OF COMPONENT C1(n).

Figure 10   Polarimetric Detection Experiment (from [3])
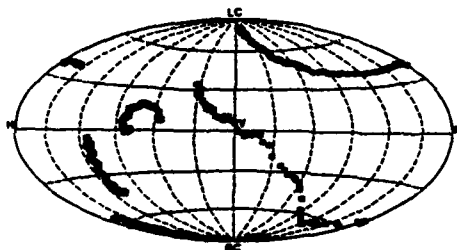
Figure 11   Time Varying Polarization States,
Five Cells of a Forest Scene (from [8])

TPR $\bar{r}_l$ of simplified aircraft model at 0° azimuth, 0° elevation.   TPR $\bar{r}_l$ of Concorde at 0° azimuth, 0° elevation.
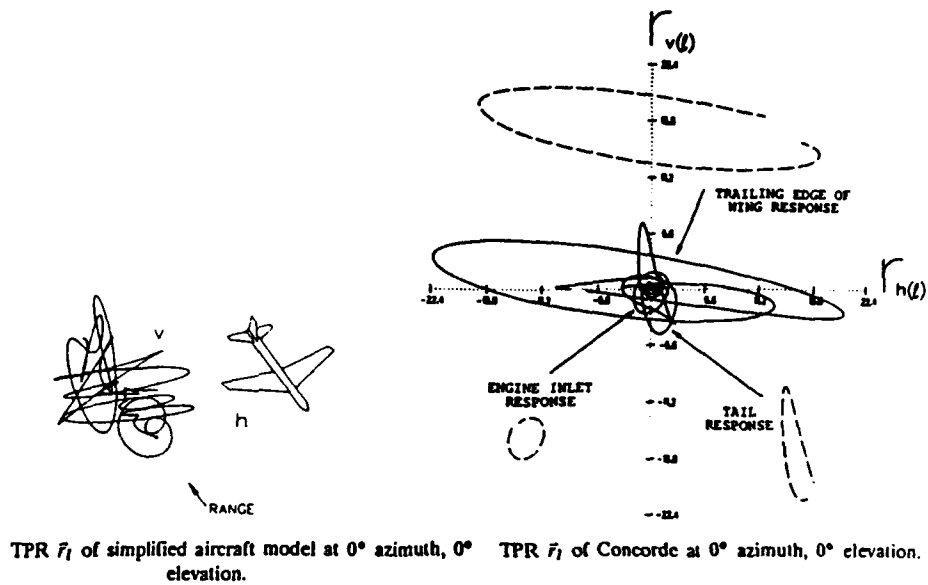
Figure 12   Transient Polarization Responses (from [9])

Data include ellipticity,
a "local" polarization angle,
and amplitude. Symbols for
down range order, here 1-4.

A "Polarization Chart"
    (this projection of the
    Poincare Sphere)

PIXEL INTENSITY DISTRIBUTIONS
POWERLINE TOWER AREA
1 ft × 1 ft RESOLUTION

Two targets shown: ◇ , ◆

◇ denotes lower hemisph

○ 1
□ 2
◇ 3
△ 4
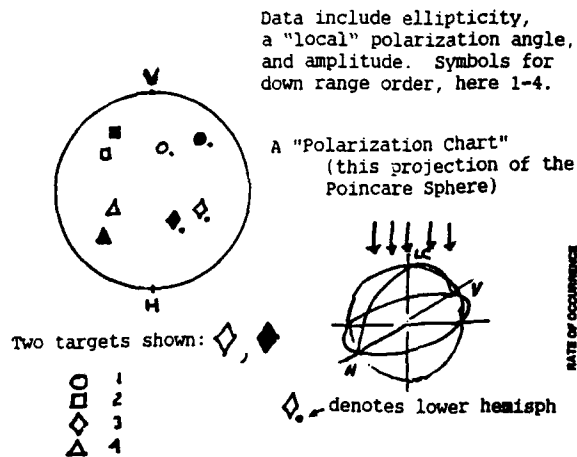
Figure 13   Dominant Scatterers, on a
"Polarization Chart" (adapted
from [10])
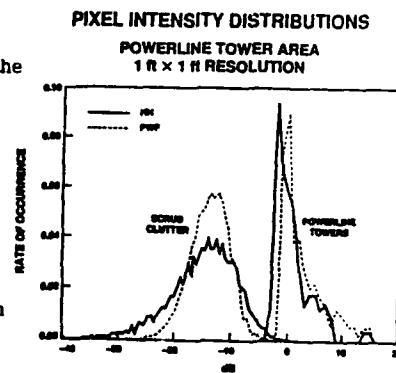
Figure 14   Histograms Showing
Effects of PWF (from [14])

BASIC THEORY OF RADAR POLARIMETRY - AN ENGINEERING APPROACH

Zbigniew H. Czyż[1]

INTRODUCTION

For engineering applications it is very important to use vectors and operators in well defined bases. Also the change of basis should be unambiguously described. Where it is possible, geometrical models of operators should be build up to clearly and faultlessly exhibit their performance. It is especially important in spinors domain, called here polarization and phase (PP) vectors which were not yet met to often in everyday engineering practice.

·Up to now, several useful definitions of polarimetric quantities have been proposed, also by this author (see [1]-[6]), but still good engineering notation for them is required.

In this paper the effort has been undertaken to meet these engineerirg needs and in the author's opinion a consistent notation, possibly most simple, precise and unambiguous has been presented.

For that purpose, earlier concepts of conjugate space necessary in polarization helix approach to the theory of radar polarimetry, or concept of polarization and phase sphere of tangent PP vectors with its interpretation as two-folded Riemann surface, or PP sphere models of bistatic in general scattering matrices, now have been stated more precisely by introducing additional notation specifying well defined orthogonal null-phase tangent vector polarization bases for vector components and matrix elements.

1. POLARIZATION HELIX /PH/ AND POLARIZATION ELLIPSE /PE/ APPROACH

In the PH approach, the polarization and phase /PP/ unit $\underline{u}$ vector is being used to describe the electric field of two plane waves, of the same polarization and phase, propagating along Oz axis of the Oxyz rectangular coordinate system in two opposite directions:

$$\underline{E}^+(t,z) = \underline{u}\, e^{j(\omega t - kz)} \tag{1a}$$

$$\underline{E}^-(t,z) = \underline{u}^*\, e^{j(\omega t + kz)} \tag{1b}$$

In the above equations $\underline{u}$ and $\underline{u}^*$ denote also the unit complex amplitudes /CA/ of the two waves, respectively. The $\underline{u}$ vector is an element of the 2-dim complex space C^2.

In the alternative, PE approach, one unit vector $\underline{u}_e$ serves as the CA vector of the two waves:

$$\underline{E}^\pm(t,z) = \underline{u}_e\, e^{j(\omega t \mp kz)} \tag{2}$$

---

[1] Telecommunications Research Institute, Warsaw, Poland

There is one-to-one correspondence: between the CA vector and the polarization ellipse, as well as between the PP vector and the polarization helix. Such helix, moving without rolling along the Oz axis in both directions, presents picturesque model of the two waves of the same polarization. And the just mentioned mutual relationships justify the names of the two approaches.

The PE for z = 0, and the PH for t = 0, both in terms of the PP vector $\underline{u}$, can be expressed by the following equations:

PE:

$$r_e^+(t) = Re\ \underline{E}^+(t,0) = Re\ \left\{\underline{u}\ e^{j\omega t}\right\} \tag{3a}$$

$$r_e^-(t) = Re\ \underline{E}^-(t,0) = Re\ \left\{\underline{u}^*\ e^{j\omega t}\right\} = r_e^+(-t) \tag{3b}$$

PH:

$$r_h(z) = z\underline{1}_z + Re\ \underline{E}^\pm(0,z)\ ;$$

$$Re\underline{E}^\pm(0,z) = Re\ \left\{\underline{u}\ e^{-jkz}\right\} = Re\ \left\{\underline{u}^*\ e^{+jkz}\right\} \tag{4}$$

It is seen from the above equations that, for identically polarized waves, the sense of the PE depends on the direction of wave propagation, while the sense of the PH does not. That is why the PH only defines unambiguously the wave polarization and phase, and the PP vector only, not the CA vector, ought to be displayed on the Poincarè polarization sphere.

This is an important conclusion justifying the PH approach to the theory of polarimetry proposed by this author in numeral publications, e.g. [1], [2] or [3], as the only correct approach.

Its advantages are immediately seen when considering for example transmissions, especially with changing polarization bases.

What only requires an explanation is the phase of the PP $\underline{u}$ vector. E.g., its phase lag means a shift of the polarization helix backwards against positive direction of the Oz axis independently of the direction of wave propagation. In time domain that means the change of phase retardation for phase advance when the wave reverses its direction of propagation. That can be seen in Fig. 1 which shows the right hand sensed PE traced in the z = 0 plane by, travelling in the positive Oz direction, a left hand sensed PH corresponding to the right elliptical polarization. After reversal of wave propagation the phase angles shown for the t = 0 point on the ellipse remain unaltered. Only their meaning will change together with the sense of the ellipse.

Two sets of parameters presented in Fig. 1: γ, δ, ε and α, β, χ will be called analytical and geometrical parameters, respectively, of the PP vector. Of course, geometrical parameters loose their meaning in other than linear orthogonal polarization basis.
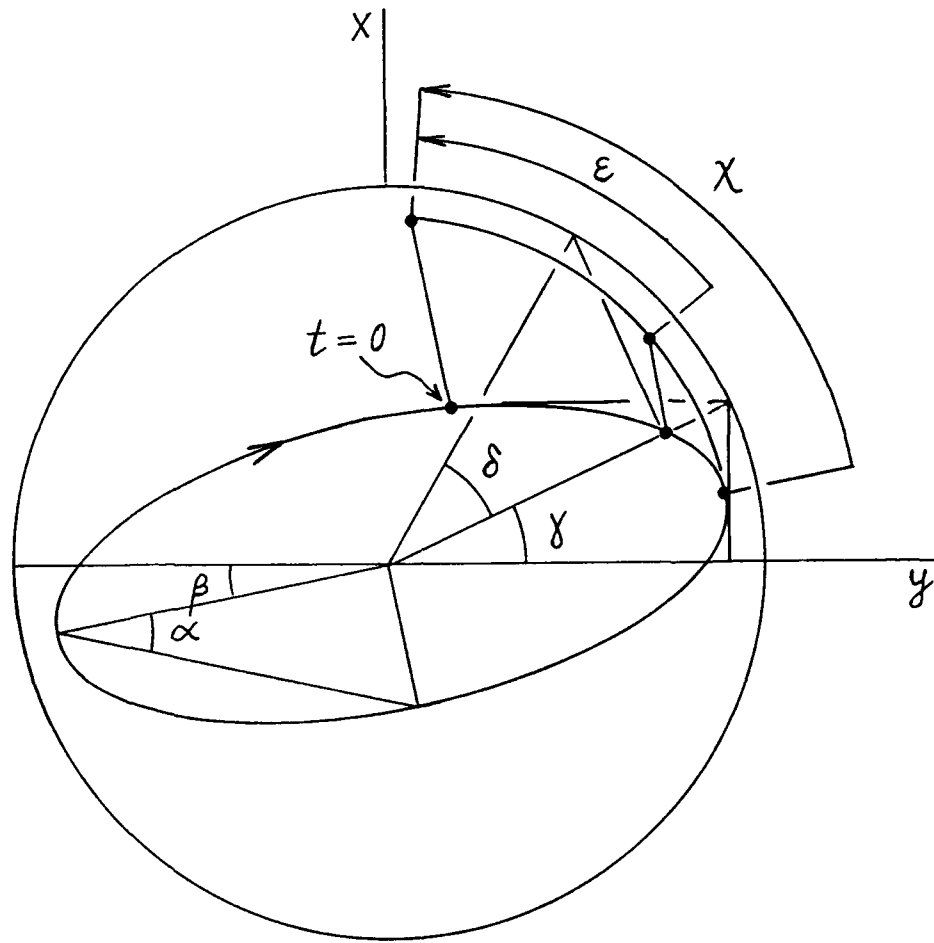
Fig. 1. Angular parameters of an oriented polarization ellipse

Fig. 2. Analytical and geometrical parameters of the PP $\underline{u}$ vector

Fig. 3. To the Hermitian multiplication of two PP vectors
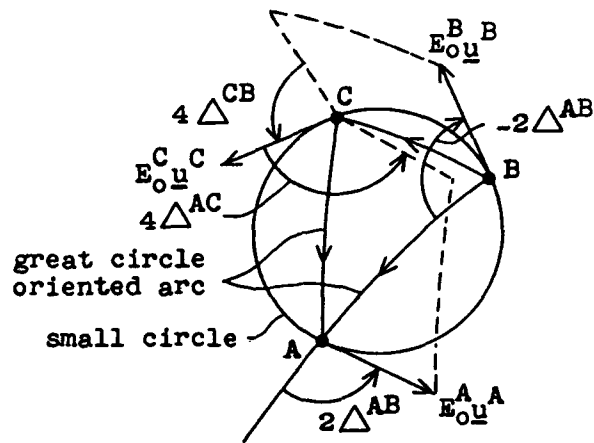
Fig. 4. Addition of two PP vectors
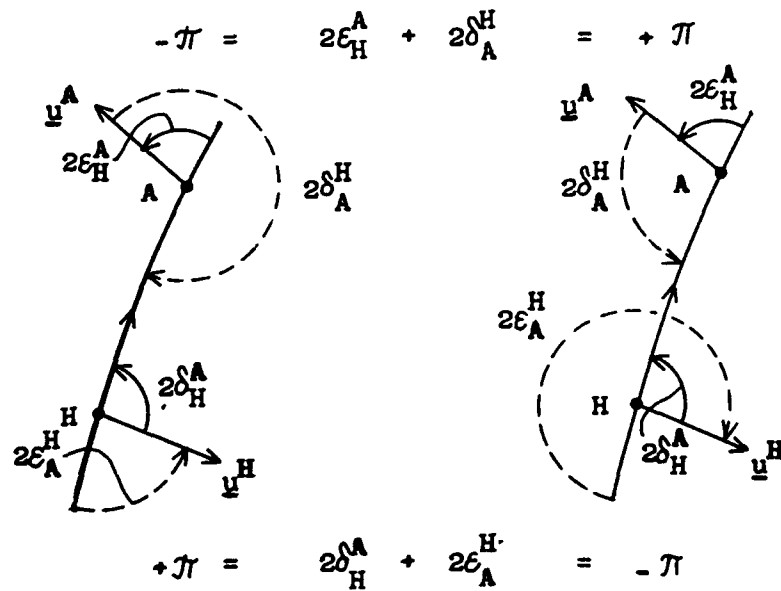
Fig. 5. Two equivalent sets of angular polarization parameters

2. POLARIZATION AND PHASE SPHERE OF TANGENT VECTORS

The conventional Poincarè sphere can be extended from 2 to 3 dimensions by introducing an additional information concerning the phase. And so, both polarization and phase can be shown by the use of vectors tangent to the sphere in their polarization points and oriented in different directions according to their phases. All angular parameters, those of polarization and, necessarily, those of phase, are doubled on the sphere, as shown in Fig 2.

Worth noticing are mutual dependences between analytical and geometrical parameters:

$$\sin 2\alpha = \sin 2\gamma \, \sin 2\delta, \qquad \cos 2\gamma = \cos 2\alpha \, \cos 2\beta \tag{5a}$$

$$\tan 2\beta = \tan 2\gamma \, \cos 2\delta, \qquad \tan 2\delta = \tan 2\alpha \, \operatorname{cosec} 2\beta \tag{5b}$$

$$\tan 2 \, (\chi - \varepsilon) = \sin 2\alpha \, \cot 2\beta = \cos 2\gamma \, \tan 2\delta \tag{5c}$$

$$\tan [\delta - (\chi - \varepsilon)] = \tan \alpha \, \tan \beta \tag{5d}$$

$$\tan [\delta + (\chi - \varepsilon)] = \tan \alpha \, \cot \beta \tag{5e}$$

In terms of those parameters the column matrix of the unit PP vector in the orthogonal linear null-phase basis $(1_y, 1_x)$ is

$$u = \begin{bmatrix} \cos\gamma & e^{-j\delta} \\ \sin\gamma & e^{+j\delta} \end{bmatrix} e^{-j\varepsilon} = \begin{bmatrix} \cos\alpha \cos\beta - j\sin\alpha \sin\beta \\ \cos\alpha \sin\beta + j\sin\alpha \cos\beta \end{bmatrix} e^{-j\chi} \tag{6}$$

This is the so called Jones column vector in its new, symmetrical form, offering numerous advantages discussed in [4].

Because of PP vector direction dependence on vector's phase, $\underline{u}$ and $-\underline{u}$ tangent at the same point of the PP sphere are identically oriented and the whole PP sphere should be considered as the two-folded Riemann surface.

Extension Pancharatnam's work [7] to tangent vectors gives natural rules of their multiplication and addition, presented in [5], with end results recalled beneath, also in Fig. 3 and Fig. 4.

For :

$$\underline{u}_1^B = \underline{u}^B \, e^{-j2\Delta^{AB}} \tag{7}$$

we get the Hermitian product

$$\underline{u}^A \cdot \underline{u}^{B*} = \underbrace{\underline{u}^A \cdot \underline{u}_1^{B*}}_{\cos \frac{AB}{2}} e^{-j2\Delta^{AB}} \tag{8}$$

with

$$2\Delta^{AB} = \varepsilon^A - \varepsilon^B + \tan^{-1}\left[\frac{\cos(\gamma^A+\gamma^B)}{\cos(\gamma^A-\gamma^B)}\tan(\delta^A - \delta^B)\right] \tag{9a}$$

$$= \chi^A - \chi^B + \tan^{-1}\left[\frac{\sin(\alpha^A+\alpha^B)}{\cos(\alpha^A-\alpha^B)}\tan(\beta^A - \beta^B)\right] \tag{9b}$$

Similarly, for:

$$\underline{u}_1^C = \underline{u}^C e^{-j2\Delta^{AC}}, \quad \text{and} \quad \underline{u}_2^C = \underline{u}^C e^{+j2\Delta^{AC}} \tag{10}$$

and for magnitude of tangent vector denoted by E_0 we get the sum

$$E_0^A \underline{u}^A + E_0^B \underline{u}^B = \left[E_0^A \underbrace{\underline{u}^A \cdot \underline{u}_1^{C*}}_{\cos\frac{AC}{2}} e^{-j2\Delta^{AC}} + E_0^B \underbrace{\underline{u}^B \cdot \underline{u}_2^{C*}}_{\cos\frac{BC}{2}} e^{+j2\Delta^{CB}}\right]\underline{u}^C$$

$$= E_0^C \underline{u}^C \tag{11}$$

with 2Δ according to equations (9), and with

$$\frac{\sin\dfrac{AC}{2}}{\sin\dfrac{BC}{2}} = \frac{E_0^B}{E_0^A} \tag{12}$$

locating C point on the Fig. 4 small circle arc.

What should be observed is that in case of symmetric orientation of $\underline{u}^A$ and $\underline{u}^B$ vectors with respect to the great circle arc AB as in Fig. 4, what always can be achieved multiplying both vectors by proper phase term, all three vectors: $\underline{u}^A$, $\underline{u}^B$ and $\underline{u}^C$ are tangent to the same small circle through points A, B, C, and are oriented in the same direction along that circle.

As shown in Fig 4, the great circle arcs BC, CA and BA are treated as oriented arcs. The orientation tells that the second vector, when looking along the arc direction, exhibits phase lag, $2\Delta > 0$, over the first one. Of course, such orientation does not depend on basis reversal discussed later, as in Fig 5.

If both vectors being added exhibit no phase difference, their sum will be on the BA arc and directed similarly as component vectors. All three vectors will reverse their direction after simultaneous change of their phases by plus or minus $\pi/2$. If the two component vectors will change their orientation by π in opposite directions, then their sum, remaining on the same great circle, will appear outside the BA arc.

3. ELLIPTIC ORTHOGONAL NULL-PHASE TANGENT VECTOR POLARIZATION BASIS

Consider the length of the AB great circle arc equal to π, denoting then the point B by Ax. The row matrix $[\underline{u}^A \ \underline{u}^{Ax}]$ of the two colinear and of the same direction vectors exibiting no phase difference, $2\Delta^{AAx} = 0$, will be called an elliptic, in general, orthogonal null-phase polarization basis. Such basis is called "null-phase" because after multiplication of its both colinear vectors by the same phase factor, they usually become not colinear.

Worth noticing is that to specify the polarization basis it is sufficient to take its first vector only, here $\underline{u}^A$, because its secord vectoru $\underline{u}^{Ax}$ is unambiguously defined by applying the proper orthogonality transformation. Note also that $\underline{u}^{Axx} = -\underline{u}^A$.

Such orthogonality transformation is often considered as one characterizing the space of PP vectors, called also spinors. It does not depend on the change of polarization basis.

4. AN ENGINEERING APPROACH - ROTATION TRANSFORMATION

To adopt general vector equations of radar polarimetry to engineering applications, the matrix form of tangent PP vectors and their operators will be used together with indication of their polarization basis, always orthogonal and null-phase one according to formerly stated definitions. And to have consistent designations of all PP vectors, $\underline{u}^H$ instead of $\underline{l}_y$ and $\underline{u}^{Hx}$ instead of $\underline{l}_x$ will be introduced with H denoting the horizontal linear polarization point on the PP sphere, and later denoting also other polarizations.

In the proposed engineering approach any $\underline{u}^A$ vector and the orthogonal $\underline{u}^{Ax}$ vector will be presented as follows:

$$\underline{u}^A = \left[\underline{u}^H \ \underline{u}^{Hx}\right] u_H^A \tag{13a}$$

$$\underline{u}^{Ax} = \left[\underline{u}^H \ \underline{u}^{Hx}\right] u_H^{Ax} \tag{13b}$$

where the row matrix $[\underline{u}^H \underline{u}^{Hx}]$ will be called the $\underline{u}^H$ polarization basis, and u_H^A and u_H^{Ax} denote column matrices of the $\underline{u}^A$ and $\underline{u}^{Ax}$ vectors in the $\underline{u}^H$ basis. According to (6) we put

$$u_H^A = \begin{bmatrix} \cos\gamma_H^A & e^{-j(\delta_H^A + \epsilon_H^A)} \\ \sin\gamma_H^A & e^{+j(\delta_H^A - \epsilon_H^A)} \end{bmatrix} = \begin{bmatrix} a_H^A \\ b_H^A \end{bmatrix} \tag{14a}$$

where a_H^A and b_H^A can be recognized as known from the literature Cayley-Klein rotation parameters. The column matrix of the orthogonal vector in the same basis can be obtained by the use of the following orthogonality transformation with the orthogonality matrix C^x:

$$u_H^{Ax} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} u_H^{A*} = \begin{bmatrix} -b_H^{A*} \\ a_H^{A*} \end{bmatrix} \qquad (14b)$$

$$\underbrace{\phantom{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}}}_{C^x}$$

It should be observed here that the orthogonality operator acts on the PP vector moved into complex conjugate space.

Any other PP vector, say $\underline{u}^B$, can expressed by column matrices in different polarization bases, say $\underline{u}^H$ or $\underline{u}^A$. Using (13a,b) we get successively :

$$\begin{aligned}
\underline{u}^B &= \begin{bmatrix} \underline{u}^H & \underline{u}^{Hx} \end{bmatrix} u_H^B = \begin{bmatrix} \underline{u}^A & \underline{u}^{Ax} \end{bmatrix} u_A^B \\
&= \begin{bmatrix} \underline{u}^H & \underline{u}^{Hx} \end{bmatrix} \begin{bmatrix} u_H^A & u_H^{Ax} \end{bmatrix} u_A^B = \begin{bmatrix} \underline{u}^H & \underline{u}^{Hx} \end{bmatrix} C_H^A u_A^B \\
&= \begin{bmatrix} \underline{u}^H & \underline{u}^{Hx} \end{bmatrix} C_H^A \widetilde{C_H^{A*}} u_H^B
\end{aligned} \qquad (15)$$

So, what has been obtained above is the definition of the rotation matrix:

$$C_H^A = \begin{bmatrix} u_H^A & u_H^{Ax} \end{bmatrix} \qquad (16a)$$

which is unitary (compare (14a) and (14b)):

$$C_H^A \widetilde{C_H^{A*}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad (16b)$$

and unimodular :

$$\det C_H^A = +1 \ , \qquad (16c)$$

the basis transformation or basis rotation rule:

$$\begin{bmatrix} \underline{u}^A & \underline{u}^{Ax} \end{bmatrix} = \begin{bmatrix} \underline{u}^H & \underline{u}^{Hx} \end{bmatrix} C_H^A \qquad (17)$$

and the change of basis rule for the PP vectors, called also their passive transformation:

$$u_A^B = \widetilde{C_H^{A*}} u_H^B \qquad (18)$$

To have the engineering approach legible enough for engineers the converse rotation matrix to that in (16a) should be

$$C_A^H = \begin{bmatrix} u_A^H & u_A^{Hx} \end{bmatrix} \qquad (19)$$

But, taking $u_H^H = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ instead of u_H^B in (18) we get

$$u_A^H = \widetilde{C_H^{A*}}\, u_H^H = \begin{bmatrix} u_A^H\ u_A^{Hx} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = C_A^H\, u_H^H$$

There from it follows

$$\widetilde{C_H^{A*}} = C_A^H \quad or \quad C_H^A\, C_A^H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{20}$$

and finally we obtain more symmetric form of the change of basis rule :

$$u_A^B = C_A^H\, u_H^B \tag{21}$$

Of course, equalities (20) impose certain relationships between angular parameters of column matrices u_H^A and u_A^H, namely:

$$2\varepsilon_H^A + 2\delta_A^H = -2\delta_H^A - 2\varepsilon_A^H = \pm\pi, \quad with \quad 2\gamma_H^A = 2\gamma_A^H \tag{22}$$

for all 2γ angles changing from 0 to 4π, and all 2δ and 2ε angles changing from -4π to $+4\pi$. Those relationships are explained in Fig 5.

Consider now the active transformation, denoting by $C_{A,K}^H$ the C_A^H matrix in a $\underline{u}^K$ basis. By that transformation we obtain from $\underline{u}^A$ a new vector $\underline{u}^H$ in the $\underline{u}^K$ basis:

$$u_K^H = C_{A,K}^H\, u_K^A = C_K^H\, C_A^K\, u_K^A = C_K^H \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{23}$$

where the rule (21) has been used. From (23) we see that

$$C_{A,K}^H = C_K^H\, C_A^K \tag{24}$$

what is interesting to compare with another equality resulting from (21):

$$C_A^H = C_A^K\, C_K^H \tag{25}$$

From (24) we see at once that

$$C_{A,H}^H = C_{A,A}^H = C_A^H \tag{26}$$

Combining (24) and (25) we obtain the change of basis transformation for rotation matrices

$$C_{A,B}^H = C_B^K\, C_{A,K}^H\, C_K^B \tag{27}$$

To gain deeper insight into the action of the rotation matrix one should observe that for P and P' points equidistant from a B point the equation

Fig. 6. Polarization and phase sphere rotation about OB axis by 2ϕ angle

Fig. 7. Polarization and phase sphere inversion through point I

$$u_B^{P'} = C_{P,B}^{P'} u_B^P = \begin{bmatrix} e^{-j\phi} & 0 \\ 0 & e^{+j\phi} \end{bmatrix} u_B^P \tag{28}$$

means rotation of the $\underline{u}^P$ vector to the $\underline{u}^{P'}$ position about an OB axis by a 2ϕ angle shown in Fig. 6. To check this, one can use (14a) to see that the only changing angular parameter of the rotated vector is

$$2\delta_B^{P'} = 2\delta_B^P + 2\phi \tag{29}$$

The equation (28) is in the $\underline{u}^B$ basis. In any other $\underline{u}^H$ basis we get

$$C_{P,H}^{P'} = C_H^B \begin{bmatrix} e^{-j\phi} & 0 \\ 0 & e^{+j\phi} \end{bmatrix} C_B^H \tag{30a}$$

$$= \begin{bmatrix} \cos\phi - j\, n_1 \sin\phi & (-n_3 - jn_2)\sin\phi \\ (n_3 - j\, n_2)\sin\phi & \cos\phi + j\, n_1 \sin\phi \end{bmatrix} \tag{30b}$$

$$= C(\underline{n}, 2\phi) = -C(\underline{n}, 2\phi + \pi) \tag{30c}$$

where the rotation unit vector $\underline{n}$ is directed along OB radius. Its components in the $\underline{u}^H$ basis are

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \begin{bmatrix} \cos 2\gamma_H^B \\ \sin 2\gamma_H^B & \cos 2\delta_H^B \\ \sin 2\gamma_H^B & \sin 2\delta_H^B \end{bmatrix} \tag{31}$$

If we want to know to what position, $\underline{u}^K$, the basis vector $\underline{u}^H$ will be moved by this rotation, we have to inspect the equality

$$C(\underline{n}, 2\phi) = C_H^K \tag{32}$$

from which we get

$$\cos\phi = \cos\gamma_H^K \ \cos(\delta_H^K + \varepsilon_H^K)$$

$$n_1 \sin\phi = \cos\gamma_H^K \ \sin(\delta_H^K + \varepsilon_H^K)$$

$$n_2 \sin\phi = -\sin\gamma_H^K \ \sin(\delta_H^K - \varepsilon_H^K) \tag{33}$$

$$n_3 \sin\phi = \sin\gamma_H^K \ \cos(\delta_H^K - \varepsilon_H^K)$$

The same equations (33) can serve to find the $\underline{n}$ axis and 2ϕ rotation angle for given the $\underline{u}^K$ vector in a $\underline{u}^H$ basis.

Any rotation matrix C_H^K can now be decomposed into product of three matrices of the form (30c) presenting vector rotation by the three Euler angles, e.g.:

$$C_H^K = C(1_{Q_H}, 2\delta_H^K) \ C(1_{V_H}, 2\gamma_H^K) \ C(1_{Q_H}, 2\varepsilon_H^K) \tag{34}$$

$$= \begin{bmatrix} e^{-j\delta_H^K} & 0 \\ 0 & e^{+j\delta_H^K} \end{bmatrix} \begin{bmatrix} \cos\gamma_H^K & -\sin\gamma_H^K \\ \sin\gamma_H^K & \cos\gamma_H^K \end{bmatrix} \begin{bmatrix} e^{-j\varepsilon_H^K} & 0 \\ 0 & e^{+j\varepsilon_H^K} \end{bmatrix}$$

$$= \begin{bmatrix} u_H^K & u_H^{Kx} \end{bmatrix}$$

where the $0 \ Q_H \ U_H \ V_H$ coordinate system of Stokes parameters corresponds to the $[\underline{u}^H \ \underline{u}^{Hx}]$, or simply $\underline{u}^H$ polarization basis, as shown in Fig 6.

To be more precise we may denote $\underline{n}$ as $\underline{n}^B$ and column vector (31) as n_H^B thus obtaining

$$\underline{n}^B = \begin{bmatrix} 1_{Q_H} & 1_{U_H} & 1_{V_H} \end{bmatrix} n_H^B \tag{35}$$

and noting the following relationship between the unit Stokes vector $\underline{P}^B$, corresponding to $\underline{u}^B$, and $\underline{n}^B$ vector, all in the $\underline{u}^H$ basis:

$$P_H^B = \widetilde{U^*}\left(u_H^B \otimes u_H^{B*}\right) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ n_H^B \end{bmatrix} \tag{36}$$

with

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & j & 1 & -j \\ 0 & 0 & 1 & j \\ 1 & -1 & 0 & 0 \end{bmatrix} ; \quad U^{-1} = \widetilde{U^*}, \quad \det U = -j \tag{37}$$

and $\otimes$ meaning Kronecker multiplication.

5. OTHER TRANSFORMATIONS IN THE ENGINEERING NOTATION

Two basic transformations of the PP vectors are rotation and inversion. The rotation has been discussed in the previous section together with the orthogonality transformation, just mentioned only because the orthogonality is a special case of inversion.
Namely, it is an inverse transformation to inversion through the center of the PP sphere.

The essential difference between the two basic transformations is that rotation leaves vectors in the same complex space, while inversion brings the vector to the conjugate space, what in physical interpretation is connected with changing of wave propagation from positive to negative direction along the Oz axis and is required by the PH approach /see Section 1/.

The inversion of the PP sphere on itself can be performed through the inversion point I inside the sphere, also in its center or on its surface. If rotation depends on 3 real parameters, e.g. 3 Euler angles or one angle 2ϕ and two angles of $\underline{n}$ vector orientation, then inversion depends on another 3 real parameters, e.g. the coordinates of the I point inside the PP sphere of unit radius.

Fundamental in radar polarimetry is the scattering transformation employing both inversion and rotation. Using the PH approach in our engineering notation the scattering or two-way transmission equation will be written for the received voltage, V_r, showing also the change of polarization basis in the next step, according to (21) and (20):

$$V_r = \widetilde{u_H^R} \; A_H \; u_H^T = \lambda^T \; \widetilde{u_H^R} \; u_H^{S*} \tag{38a}$$

$$= \widetilde{u_H^R} \; \widetilde{C_K^H} \; \widetilde{C_H^K} \; A_H \; C_H^K \; C_K^H \; u_H^T \tag{38b}$$

$$= \widetilde{u_K^R} \; A_K \; u_K^T = \lambda^T \; \widetilde{u_K^R} \; u_K^{S*} \tag{38c}$$

As we see, change of basis coserves the equation form in the PP approach.

The amplitude scattering matrix A in the, say, $\underline{u}^H$ basis can be decomposed into following factors:

$$A_H = \frac{\sqrt{\sigma_0}}{2} \; C_{P,H}^{K*} \; A_H^{INV} \; e^{j\xi} \tag{39}$$

Two of them are independent of the basis, so A without lower index H will be used in their description. They are: magnitude factor $\sqrt{\sigma_0}/2$, being radius of the PP sphere model of A, with

$$\sigma_0 = Span\ A + 2\ |\det A| \tag{40}$$

and phase factor, argument of which is

$$\xi = \frac{1}{2}\ arg\ \det A \tag{41}$$

Conjugate rotation matrix is used in (39) because it operates after inversion in the conjugate space. K and P indices recall that in the $\underline{u}^K$ basis, characteristic to the matrix A, a P point, moving by the rotation to K, has special meaning for the A matrix geometrical model. Namely, the AB axis of rotation after inversion, which is in the $Q_K = 0$ plane, can be defined by $n_{2K}^B = \sin 2\delta_K^P$ and $n_{3K}^B = -\cos 2\delta_K^P$.

Applying the way of numbering of A matrix elements after van de Hulst [9]

$$A_H = \begin{bmatrix} A_{2H} & A_{3H} \\ A_{4H} & A_{1H} \end{bmatrix} \tag{42}$$

we get, as in [4]:

$$C_{P,H}^{K\ *} = \frac{1}{\sqrt{\sigma_0}} \begin{bmatrix} A_{3H}e^{-j\xi} - A_{4H}^{*}e^{j\xi} & -A_{2H}e^{-j\xi} - A_{1H}^{*}e^{j\xi} \\ A_{1H}e^{-j\xi} + A_{2H}^{*}e^{j\xi} & -A_{4H}e^{-j\xi} + A_{3H}^{*}e^{j\xi} \end{bmatrix} \tag{43}$$

The n_H^B axis and 2ϕ angle of rotation can now be found easily by comparing the above matrix with conjugate form of (30b).

The inversion matrix, expressed also by A_H matrix elements, will next be given a geometrical interpretation as in case of rotation transformation. With $M_{iH} = A_{iH}\ A_{iH}^{*}$ we get

$$A_H^{INV} = \frac{2}{\sigma_0} \begin{bmatrix} A_{2H}A_{3H}^* + A_{4H}A_{1H}^* & M_{3H} + M_{1H} + |\det A| \\ -M_{2H} - M_{4H} - |\det A| & -A_{3H}A_{2H}^* - A_{1H}A_{4H}^* \end{bmatrix}$$

$$= \begin{bmatrix} (-m_2 - jm_3)\cos 2\Gamma & 1 + m_1 \cos 2\Gamma \\ -1 + m_1 \cos 2\Gamma & (m_2 - jm_3)\cos 2\Gamma \end{bmatrix}$$

$$= A^{INV}(\underline{m}, 2\Gamma) \tag{44}$$

$$= \begin{bmatrix} -U_H^I - jV_H^I & 1 + Q_H^I \\ -1 + Q_H^I & U_H^I - jV_H^I \end{bmatrix}$$

On the PP sphere of unit radius, for given inversion point I, are shown in Fig. 7: "angle of inversion" 2Γ, "inversion axis" $\underline{m} = \underline{m}^N$ indicating the polarization point of minimum scattered power, and easy to check rule of phase transformation while inversion.

It is worth noticing that components of the $\underline{m}^N$ vector in a $\underline{u}^H$ basis, which are proportional to I point components:

$$m_H^N = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} Q_H^I \\ U_H^I \\ V_H^I \end{bmatrix} \sec 2\Gamma = \frac{-2}{\sigma_0} \begin{bmatrix} b_{1H} \\ b_{3H} \\ b_{5H} \end{bmatrix} \tag{45}$$

can be directly expressed by 3 elements of the first row of Kennaugh (Stokes) matrix

$$F_H = \widetilde{U}\left(A_H \otimes A_H^*\right) U = \begin{bmatrix} a_{1H} & b_{1H} & b_{3H} & b_{5H} \\ c_{1H} & a_{2H} & b_{4H} & b_{6H} \\ c_{3H} & c_{4H} & a_{3H} & b_{2H} \\ c_{5H} & c_{6H} & c_{2H} & a_{4H} \end{bmatrix} \tag{46}$$

Notation proposed here for matrix elements is after Perrin [8] and has been adopted also by van de Hulst [9].

The change of basis transformation is conceptionally simpler in four dimensions because it uses only orthogonal rotation matrix

$$D_H^A = \widetilde{U^*}\left(C_H^A \otimes C_H^{A*}\right) U = \widetilde{D_A^H} \tag{47}$$

but still rotation by 3 Euler angles remains.

In terms of one axis and one angle of rotation the D matrix takes the form

$$D(\underline{n}, 2\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2\phi + \left(n_1^2 - n_2^2 - n_3^2\right)\sin^2\phi & -n_3 \sin2\phi + 2n_1 n_2 \sin^2\phi & n_2 \sin2\phi + 2n_1 n_3 \sin^2\phi \\ 0 & n_3 \sin2\phi + 2n_1 n_2 \sin^2\phi & \cos^2\phi + \left(-n_1^2 + n_2^2 - n_3^2\right)\sin^2\phi & -n_1 \sin2\phi + 2n_2 n_3 \sin^2\phi \\ 0 & -n_2 \sin2\phi + 2n_1 n_3 \sin^2\phi & n_1 \sin2\phi + 2n_2 n_3 \sin^2\phi & \cos^2\phi + \left(-n_1^2 - n_2^2 + n_3^2\right)\sin^2\phi \end{bmatrix} \tag{48}$$

CONCLUDING REMARKS

Evidently, it was not possible in short paper to fully develop the complete engineering notation for radar polarimetry. Many fundamental problems as e.g. one-way transmission considered as scattering by the free space, Lorentz transformation as a sequence of inversion and orthogonality, or scattering by partially depolarizing medium and many, many other have not been even mentioned. Just few ideas have been presented for readers approval.
Among them most important in author's opinion are:

1° the use of null-phase orthogonal basis presented by its first tangent vector only,

2° introducing the upper and lower indices for vector and basis notation, respectively, given by polarization points only without specifying their phases, for simplicity reasons. (In case of two or more vectors tangent at the same point to the PP sphere, vectors could be distinguished by additional numbers as lower indices. See for example equalities (10). For column matrices these numbers in lower indices could be brought to the labels of points what would result in notation as e.g. $u_{C_2}^{B_1}$ which seems to be not too much complicated.)

REFERENCES

1. Z. H. Czyż: Amplitude and power representation of the elliptically polarized waves (in Polish). Prace PIT, No. 63, pp 11-22, 1969.

2. Z. H. Czyż: Polarization of radar scatterings (in Polish). Prace PIT Supplement No. 5, pp 1- 154, 1986.

3. Z. H. Czyż: Comparison of fundamental approaches to radar polarimetry. In: Direct and Inverse Methods in Radar Polarimetry, Part 1. Kluwer Academic Publishers, Editor: W. M. Boerner, pp 99-116; 1992.

4. Z. H. Czyż: Comparison of polarimetric radar theories. Proc. of the Second International Conference on Electromagnetics in Aerosparce Applications, Torino, Italy, pp 291-294, Sept. 17-20, 1991.

5. Z. H. Czyż: An alternative approach to foundations of radar polarimetry. In: Direct and Inverse Methods in Radar Polarimetry, Part 1, Kluwer Academic Publishers, Editor: W. M. Boerner, pp 247-266, 1992.

6. Z. H. Czyż: Polarization properties of nonsymmetrical matrices. IEEE Trans. on Aerospace and Electronic Systems, Vol. 27, No. 5, September, pp 771-777 and 781-782, 1991.

7. S. Pancharatnam: Generalized theory of interference and its applications, Part 1. Proceedings of Indian Academy of Sciences, Vol. 44A, pp 247-262, 1956.

8. F. Perrin: Polarization of light scattered by isotropic opalescent media. Journ. Chem. Phys., Vol 10, pp 415-427, 1942.

9. H. C. van de Hulst: Light scattering by small particles. Wiley, New York, 1957.

MICROWAVE RESEARCH IN AGRICULTURE

Andrzej W. Kraszewski[1]

Abstract. Microwave techniques, methods and instrumentation can be utilized to increase crop production and improve crop quality. Some examples of successful microwave research in agriculture are presented and future prospects are discussed.

INTRODUCTION

Farming is one of the oldest and most important human activities, and it is essential to the survival of the species. The level and effectiveness of agriculture was for centuries a determinant of the wealth and prosperity of nations, and it still is today. Therefore, it is no wonder that discoveries in all other branches of science and technology have been tried in agriculture to increase its effectiveness, to increase crop production and improve crop quality, as well as to reduce the effort and cost of crop producion. In this way, first plows and irrigation systems, then tractors and artificial fertilizers and pesticides were developed, and recent developments include satellite sensing for crop prediction and global positioning systems (GPS) for crop planning and production optimization. As electromagnetic radiation, including microwaves, plays an ever increasing role in modern life, it is quite understandable that the two recent developments in modern agriculture just mentioned are related to microwave technology applied to satellite communication. But by no means are existing and possible applications of microwave techniques and instrumentation in agriculture limited to those two examples. Different aspects of microwaves applied in various branches of science and technology have been reviewed recently [1-6]. The purpose of this paper is to present briefly recent results and future trends of microwave research in agriculture.

Unique features of microwave radiation that make it useful for industrial and agricultural applications can be listed as follows:
- microwaves can propagate through free space allowing remote sensing to be accomplished,
- they propagate along straight lines and reflect from metal surfaces, observing the laws of optics,
- solid dielectric materials are opaque to light and infrared radiation but transparent to microwaves, which permits the probing of the whole volume of objects transported inside a dielectric tubing without the need for inserting windows,
- some materials like gases and water react specifically with selected microwave frequencies, allowing measurements of small amounts of water and gas concentration in complex mixtures,
- microwave radiation does not alter nor contaminate the material under test, enabling fast, nondestructive and continuous monitoring,
- the effect of DC conductivity which decreases with frequency, is much smaller than at radiofrequencies, which makes moisture measurement easier,
- in contrast to ionizing radiation, microwave methods are much safer and very fast,
- microwave radiation is relatively insensitive to environmental conditions; thus dust and water vapor do not affect the measurement, in contrast to infrared methods.

During the past thirty years, these features of microwave radiation were utilized to solve some problems in agriculture and were tried on many others. It would be beyond the scope of this paper to mention all of them. However, the most essential applications and the most promising solutions are presented here and their future developments are discussed.

---

[1] U.S. Dept. of Agriculture, Agricultural Research Service, Richard B. Russell Agricultural Research Center, Athens, Georgia 30613, U.S.A.

GENERAL PRINCIPLES

Properties of Materials. The parameters of materials which describe their interaction with electromagnetic fields are the permittivity ϵ^* and the permeability μ^*. The permittivity describes the material behavior in the electric field and consists of a real part, ϵ', called the dielectric constant, and an imaginary part, ϵ'', called the loss factor. Thus, the permittivity is expressed as

$$\epsilon^* = \epsilon' - j\epsilon'', \tag{1}$$

where the dielectric constant represents the ability of a material to store electric energy, and the loss factor describes the loss of electric field energy in the material. Another parameter frequently used is the dissipation factor, called also the loss tangent, defined as the ratio of the loss factor to the dielectric constant. The permittivity of materials is often normalized to the permittivity of vacuum and referred to as the *relative permittivity*

$$\epsilon_r^* = \frac{\epsilon^*}{\epsilon_0} = \frac{\epsilon'}{\epsilon_0} - j\frac{\epsilon''}{\epsilon_0}, \tag{2}$$

where $\epsilon_0 = 8.854 \times 10^{-12}$ F/m. The relative dielectric constant and loss factor are thus dimensionless quantities. The relative loss factor is a function of the material conductivity

$$\epsilon_r'' = \frac{\epsilon''}{\epsilon_0} = \frac{\sigma}{2\pi f \epsilon_0} \tag{3}$$

where σ is the conductivity in S/m and f is the operating frequency in Hz. The permittivity of dielectric materials depends on operating frequency, temperature, composition (in particular water and salt content), density and physical state (solid or liquid, e.g., water and ice have dramatically different dielectric properties at microwave frequencies).

The permeability μ^* consists of the real magnetic permeability μ' and the magnetic loss factor μ''
$$\mu^* = \mu' - j\mu'',$$
and like the complex permittivity, the permeability can be normalized to that of vacuum:

$$\mu_r^* = \frac{\mu^*}{\mu_0} = \frac{\mu'}{\mu_0} - j\frac{\mu''}{\mu_0} \tag{4}$$

where $\mu_0 = 1.26 \times 10^{-6}$ H/m is the magnetic permeability of vacuum. The real part of the magnetic permeability represents the ability of the material to store magnetic field energy, while the magnetic loss factor describes the loss of magnetic field energy in the material.

The dielectric properties of agricultural materials, which are in general hygroscopic, vary predominantly with moisture content, but also they depend on the frequency of the applied electromagnetic field, the temperature of the materials, and on density and structure of the materials. In granular or particulate materials, the bulk density of the air-particle mixture is another factor that influences their dielectric properties. The bulk density can be altered by the shape and dimensions of particles and their surface conditions. Interest in the dielectric properties of agricultural materials extends from growing vegetation in various stages of maturity, to fruits, seeds and nuts after harvest and during storage and trade, both in bulk materials and as individual kernels and nuts. Such a broad spectrum of material structure and quantities requires practically all existing microwave methods of permittivity measurements to be used in agricultural research, including time-domain reflectometry and broadband frequency-domain spectroscopy. The main difficulty in carrying out experiments with agricultural materials is their fragile structure that changes permanently with time. For example, water evaporation at room temperature, accelerated evaporation at elevated temperature, chemical processes related to plant growth and riping, all make replicating experiments with the same material virtually impossible. However, recently available broadband, computer controlled instrumentation enables very fast measurements, and modern mathematical routines provide opportunities for using such data to develop more and more precise dielectric models [7].

Plane Wave in Dielectric Medium. The propagation constant of a plane electromagnetic wave in a lossy dielectric medium is defined as:

$$\gamma = \alpha + j\beta = \sqrt{j\omega\mu^*(\sigma + j\omega\epsilon')}, \tag{5}$$

where α is the attenuation constant, β is the phase constant, and $\omega = 2\pi f$ is the angular frequency. For a nonmagnetic material, when $\mu^* = \mu_0$, the propagation constant may be expressed as

$$\gamma = j\frac{2\pi}{\lambda}\sqrt{\epsilon_r^*} \tag{6}$$

where λ is the wavelength in free space, and $f = \frac{c}{\lambda}$ where $c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$ is the speed of light. The two components of the propagation constant can now be rewritten as:

$$\alpha = \frac{2\pi}{\lambda}\sqrt{\frac{\epsilon_r'}{2}\left(\sqrt{1 + \tan^2\delta} - 1\right)} \qquad (7)$$

$$\beta = \frac{2\pi}{\lambda}\sqrt{\frac{\epsilon_r'}{2}\left(\sqrt{1 + \tan^2\delta} + 1\right)} \qquad (8)$$

The wavelength in a lossy dielectric material is then

$$\lambda_\epsilon = \lambda\sqrt{\frac{\epsilon_r'}{2}\left(\sqrt{1 + \tan^2\delta} + 1\right)}. \qquad (9)$$

When a plane wave is incident normally upon a dielectric interface as shown in Fig. 1a, part of it is reflected and part transmitted inside the material. The reflection coefficient is equal to

$$R = \frac{E_R}{E_0} = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \qquad (10)$$

where E_0 is the incident electric field vector and E_R is the reflected electric field vector, and η_1 and η_2 are the intrinsic wave impedances of dielectric media 1 and 2, respectively;

$$\eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}} \qquad \text{and} \qquad \eta_2 = \sqrt{\frac{\mu_2}{\epsilon_2}}. \qquad (11)$$

The transmission coefficient from medium 1 to 2 is expressed as

$$T = \frac{E_T}{E_0} = 1 + R = \frac{2\eta_2}{\eta_1 + \eta_2}, \qquad (12)$$

and the transmitted power is

$$P_T = P_0\left(1 - |R|^2\right) \qquad (13)$$

Fig. 1. Plane wave normally incident on:
a) semi-infinite dielectric plane;
b) a dielectric layer of thickness d.

where P_0 is the incident power. If the dielectric 1 is free space (or air), and the dielectric 2 has $\mu^* = \mu_0$, then the reflection coefficient at the interface can be written as

$$R = \frac{1 - \sqrt{\epsilon_r^*}}{1 + \sqrt{\epsilon_r^*}}. \qquad (14)$$

It is evident from this equation that both the modulus and phase angle of the reflection coefficient, as well as the transmission coefficient, depend upon the permittivity, i.e., both the dielectric constant and loss factor (or conductivity) of the material. Thus, the information about a test parameter that is a function of the permittivity can be obtained by measuring the modulus and/or phase of the reflection or the transmission coefficients. Choosing either of these combinations is a matter of careful consideration for each given application.

The reflection and transmission coefficients for a dielectric layer of thickness d and for normal wave incidence can be calculated by considering the wave and load impedances in various regions, as shown in Fig. 1b. The reflection coefficient in region 1 is

$$R_1 = \frac{Z_{L1} - \eta_1}{Z_{L1} + \eta_1}, \qquad (15)$$

where

$$Z_{L1} = \eta_2\left(\frac{\eta_3 \cos k_2 d + j\eta_2 \sin k_2 d}{\eta_2 \cos k_2 d + j\eta_3 \sin k_2 d}\right) \qquad (16)$$

and $k_2 = \frac{2\pi}{\lambda_{\epsilon 2}}$, and $\lambda_{\epsilon 2}$ is the wavelength in dielectric region 2. The ratio of the wave transmitted into region 3 to the incident wave in region 1 can be written as

$$T_3 = \frac{E_3}{E_0} = \frac{2\eta_3}{\eta_2 + \eta_3} R_r e^{-jk_2 d} \qquad (17)$$
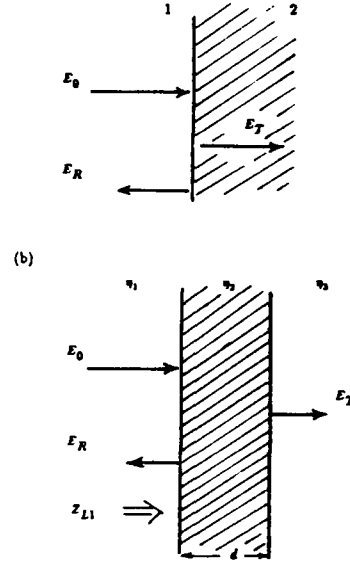
It maybe noted from Eqns. (15) - (17) that for a constant layer thickness, information about the layer permittivity can be obtained from either reflection or transmision coefficients. The reader should be aware of the fact that, in practice, radiating elements are placed relatively close to the material surface and therefore operate in the near field, and that usually, the surface of the material is not ideally smooth nor flat. This means that the incident waves are only approximations of plane waves and that the reflection and transmission coefficients depend on the position of the material with respect to the antenna. Moving the antenna away from the material surface, e.g., to ensure far-field conditions, requires a test sample of material with a large surface area, and may cause problems due to scattering from other objects in the vicinity. In any case, however, the above equations can be used to estimate the measured quantities, and usually this approximation is good enough for practical purposes.

The process of microwave heating consists of dissipating part of the microwave energy that flows into a heated material, which, in general, is a lossy dielectric. As for any heating process, the heating rate can be expressed as

$$\frac{dT}{dt} = \frac{P_d}{C\rho} , \tag{18}$$

where T is the temperature of the heated material in °C, t is time in s, P_d is the power dissipated in the heated material in W/m^3, C is the specific heat of the material in $J/kg°C$ and ρ is the density of the material in kg/m^3. The power dissipated in the material in a plane wave, when expressed in W/m^3, is

$$P_d = \frac{\omega\epsilon_0}{2} \int_V \epsilon''|E|^2 dV + \frac{1}{2} \int_V \sigma_{DC}|E|^2 dV, \tag{19}$$

where σ_{DC} is the conductivity of the material in S/m and E is the electric field intensity (peak value) in V/m. For most materials heated by microwaves, σ_{DC} is small compared with $\omega\epsilon''$, and the equation for most practical application has the form:

$$P_d = 27.8 \times 10^{-6} f \int \epsilon''|E|^2 dV, \tag{20}$$

where f is the operating frequency in MHz. To calculate the total power dissipated in a heated material at a given frequency, the distribution of the loss factor and the internal electric field intensity must be known. This is rarely true in practice, and therefore some simplifying assumptions have to be made. Most frequently, it is assumed that the loss factor does not vary appreciably within the heated volume and that the heated object is large compared with the wavelength.

For a heated body in the form of a slab occupying an infinite half space, as shown in Fig. 1a, and exposed to a plane wave at normal incidence, the incident power is partly reflected and partly transmitted and dissipated in the heated object. The distribution of the dissipated power in a homogeneous lossy material follows an exponential law, i.e.,

$$P_d = P_T e^{-2\alpha z} = P_0 \left(1 - |R|^2\right) e^{-2\alpha z} , \tag{21}$$

where α is the attenuation constant of the material given by Eqn. (7), z is the distance along the direction of wave propagation, and R is the reflection coefficient determined in Eqn. (10). For objects of dimensions comparable to the wavelength, the internal electric field in Eqn. (20) becomes a complex function of the dimensions and shape of the object, as well as the permittivity distribution inside the material.

Doppler Effect. The phase of a plane wave reflected from an interface (metal or dielectric) depends not only on the permittivity but also on the distance between a sensor (antenna) and the target (interface). When a continuous wave (CW) at constant frequency is reflected from a surface located at a distance z, and when the reflected wave is combined with the incident wave in a nonlinear element (a mixer diode), the product is

$$U_m = kE_0 \sin \frac{2\pi f \times 2z}{c}, \tag{22}$$

where E_0 is the intensity of the incident wave, k is a constant describing the mixer, and c is the velocity of light. If the target is moving relative to the antenna

$$U_m = kE_0 \sin \left(2\pi f \frac{2v}{c} t\right), \tag{23}$$

where v is the target velocity in the direction of the incident wave and t is time. The phase information in Eqn. (22) is not sufficient to determine the distance z unambiguously because of the 2π-periodicity of the phase. However, it can be used to determine the small vibrations of a fixed target or the velocity of a moving target. This is the so-called Doppler effect. When the object is moving at an angle θ with respect to the direction of the wave propagation (antenna beam), the Doppler frequency is

$$f_D = 2\,\tfrac{v}{c}\,f\,\cos\theta. \tag{24}$$

If the microwave signal is sinusoidally modulated in frequency at a relatively low frequency f_m, such a system can be used for precise distance measurements. This is known as a frequency-modulated continuous-wave (FMCW) radar and is described elsewhere [1].

APPLICATIONS OF REFLECTED WAVES

In this section, applications of microwave reflected waves for agricultural purposes are discussed, and examples are related mainly to various applications of radar.

Remote Sensing Study. Monitoring soil and vegetation biophysical status on regional and global scales is of primary importance for understanding and protecting the environment, as well as for natural resources management. The potential of remote sensing techniques, including RF, microwave, IR and visible radiation, is recognized for operational services such as crop planning and forecasting, although at present, they do not seem reliable enough and require further experimental and theoretical research. Using active microwave systems operating in the frequency range from 1 to 25 GHz provides data for classifying agricultural crops in conjunction with other observations. Active microwave responses are selective for canopy structure and the dielectric properties of the target, whereas visible and infrared responses are primarily sensitive to chlorophyll content, surface-moisture changes and soil background color [8, 9]. Current studies are using active microwave (radar) data to identify the sensitivity to near-surface soil moisture and vegetation biomass as a function of surface roughness and soil texture of agricultural fields at different phases of crop growth for various combinations of radar sensor parameters, such as frequency, polarization and angle of incidence. The other variables to be taken into considerations are: soil-moisture profile, soil bulk density and soil surface boundary conditions (random surface roughness, row direction effect related to ridge/furrow tillage practice, local slopes, etc.), vegetation canopies (variety of crop, height, leaf size, etc.), and geographical conditions, such as variability in local topography, soil texture, field size and shape, and the presence of nonagricultural features such as urban areas, forests, water bodies, etc.

The scientific rationale for conducting dielectric investigations are clearly twofold: first, to gain a fundamental understanding of the basic principles governing microwave sensor response and, second, to provide an accurate data base for the derivation of dielectric properties as needed inputs for increasingly accurate and demanding microwave scattering models.

Migration Study of Insects and Birds. For some time now, short-range radar sensors have been recognized as useful tools in ornithological and entomological research. Seasonal migration of insects and birds is of great importance for safety of growing crops; thus, tracking their routes in certain areas is crucial. The numbers of birds or insects airborne at one time can be impressive, even in the absence of social factors that cause flocking (birds) or swarming (insects). The simultaneous movement of a large number of individuals occasioned by a migratory stimulus is well distinguished by a radar sensor. Often such flights take place at night and this is when the superiority of radar sensing over visual techniques is most evident. Study of the effective radar cross-section of small living objects, improving sensor parameters (sensitivity, signal processing, etc.) and data analysis are among the most important goals of current research [10].

Monitoring Velocity of Grain Flow. The Doppler effect can be employed in monitoring the mass flow rate of grain and other particulate solids. The technique is based on measurement of the frequency and amplitude of the microwave signal scattered back by the particulates in the flow field. Essentially, two configurations are possible – monostatic
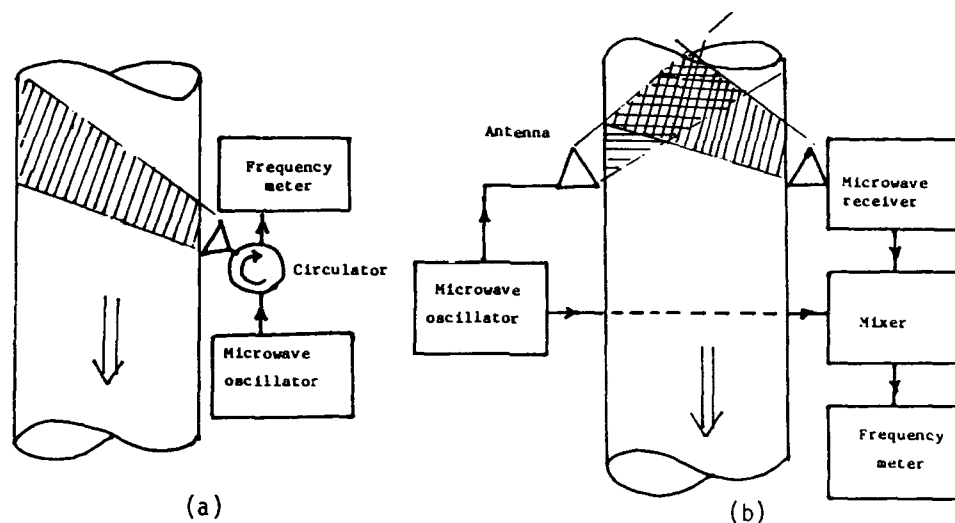
Fig. 2. Microwave Doppler velocity meters: a) monostatic configuration; b) bistatic configuration.

and bistatic, as shown in Fig. 2a and 2b, respectively. A microwave signal is directed into the flow by a horn antenna. The transmitted signal is scattered by solid particles in the flow and the reflected signal is detected by a receiver and mixed with a portion of the transmitted signal to obtain the Doppler frequency. At a typical X-band frequency of 10.5 GHz, this detected Doppler frequency is approximately 700 Hz for a 10 $m \cdot s^{-1}$ change in solids velocity. For both configurations, however, the detected Doppler signal does not consist of just one frequency component but will be composed of a number of different frequencies. This spread of frequencies is a result of a divergent antenna beam being reflected from a number of particles in the interaction area, as well as the random nature of the particle movement. However, through proper calibration and signal processing, the strong correlation between the mean Doppler frequency and the average velocity of grain was established. In further research [11] it was shown that the Doppler signal reflected from the flowing solids not only contains information about the average velocity of these particles but also that the density of the flowing material can be determined from the RMS value of the Doppler signal. Thus, a calibrated microwave Doppler radar can be used as a simple and inexpensive monitor of the velocity and the density of the flowing grain. Such a monitor provides fast response and no obstruction to the flow, since it can be installed outside a plastic pipeline used for grain transport, which does not need to be optically transparent.

A Doppler radar was successfully applied to monitor the velocity of tractors (true ground speed) and of total distance travelled with accuracies up to 0.5%. By use of two Doppler radars together, the tractor wheel slip during field tests was measured with accuracy satisfactory for practical application [2]. The distance measuring techniques using the Doppler FM radars can be used for monitoring of a material surface level in grain elevators and proximity sensors can be used for warning and protecting tractors and combines against collisions. These are typical applications of the technique discussed in [1].

APPLICATION OF TRANSMITTED WAVES

Studies of the applications of transmitted microwave radiation in agricultural research can be divided into two groups based on the level of applied power. All applications in which microwave power is used to increase temperature of the material will be listed under high power applications (heating and drying). The remaining applications are considered as low power applications (microwave instrumentation).

High-Power Applications

Heating and Drying. There are some well established applications of microwave power in agricultural product and food processing. Among them the following should be mentioned as the most successful [2, 6]:

- **meat tempering:** this refers to heating of meat frozen at $-18°C$ to a temperature of -1 or $0°C$ whereby it can be easily cut or ground to produce a variety of packaged foods such as frozen hamburger. Prior to the ability of microwave defrosting, the meat was set out in large rooms to thaw, taking several days and losing up to 3% of its juice as drip loss. Microwave tempering occurs with meat still in its cardboard containers, and takes only 5 minutes. The savings in the drip loss alone are often sufficient to pay for the installation within a year;
- **cooking bacon:** bacon for fast food outlets is precooked in central locations by 300 kW microwave cookers that process 1000 kg/hour. Precooking the bacon 80% to 90% removes most of the water and fat so that the final cooking at the fast food outlet is quick and clean. The microwave cooking eliminates shrinkage and reduces nitrosamines by heating uniformly at lower temperatures than either a grill or a broiler. Cooking at lower temperatures also enhances flavor;
- **pasta drying:** in this process microwave energy is used only to "finish-dry" pasta. Conventional hot-air drying is used to evaporate most of the moisture. Only the residual moisture deep within the pasta is evaporated with microwave power. In fact, the microwave power tends to drive some of this deep moisture to the surface where it is evaporated using low cost hot air;
- **drying crushed oranges:** crushed oranges are dried in a low vacuum using microwave power to produce a powder capable of being reconstituted into orange juice that is hardly distinguishable from fresh orange juice. The low drying temperatures protect the flavor from degradation;
- **soybean dehulling:** drying soybeans with microwave power immediately prior to dehulling at a subatmospheric pressure eliminated the need for 4 days of storage time (for moisture equilibrium) and provided superior separation after dehulling. The payout for the installation was reasonable, but its very high initial cost prevented practical use in a low-margined commodity business;
- **drying peanuts:** experimental drying of peanuts with microwaves produced some very tasty products, but the seasonal nature of the drying did not allow adequate annual usage of the expensive microwave equipment to pay for itself in a reasonable time;
- **sterilizing delicate foods:** microwaves are ideal for sterilizing heat sensitive foods after they have been placed in their containers and sealed. Because microwaves heat throughout the material, heat-up times can be very short; the outer edges of the material need not be exposed to higher temperatures than the inner material, and overall exposure of the material to high temperatures can be very short. Experiments showed that dill pickles sterilized by microwaves were much crisper and had a better texture;
- **grape puffs:** entirely new foods can be created with microwave processing; recently, grape puffs have been produced by drying grapes with microwave power in a low vacuum. The microwave energy evaporated the water and subsequently caramelized the grape sugar to produce a sweet crunchy snack. The grape retained nearly its original shape and its full color, becoming a completely new premium product;
- **aging wine and spirits:** in wine irradiated at 915 MHz, the amount of acid decreased and that of esters increased more than in six months unprocessed spirits, providing better flavor and saving storage space and equipment;
- **cacao-bean roaster:** a multimode cavity applicator was used in the form of a horizontal cylinder that rotated slowly allowing beans to be exposed to 5 kW CW power at 2.45 GHz.

Selective Heating – Controlling Stored-Grain Insects. In the study of insect control by dielectric heating, knowledge of the dielectric properties of the grain and the insects was needed. The properties of hard red winter wheat, *Triticum aestivum* (L.), and of adult insects of the rice weevil, *Sitophilus oryzae* (L.), were measured over the frequency range from 1 MHz to 12 GHz and examined for the optimum frequencies for selective dielectric heating of the insects [12]. Resulting values for the insects and the grain are shown in Fig.
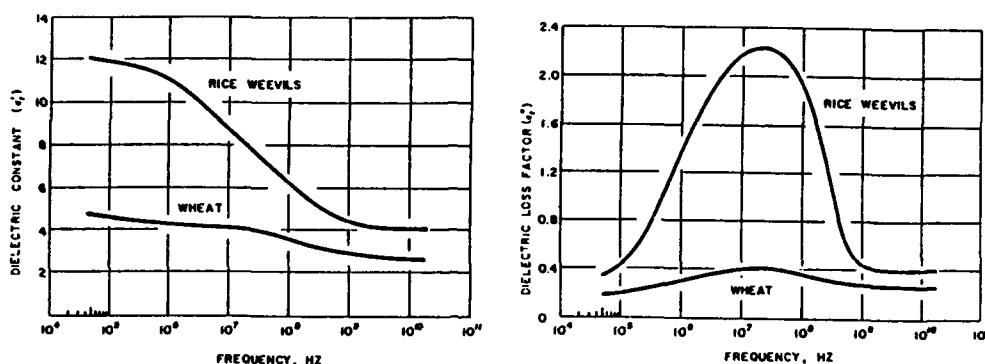
Fig. 3. Dielectric dispersion and absorption at 24°C for bulk samples of adult rice weevils and hard red winter wheat of 10.6% moisture content [12].

3, where loss-factor comparison reveals an obvious advantage for selectively heating the insects in the 10-100 MHz region. This result was confirmed by achieving complete insect mortality with much lower energy input at 39 MHz than at 2.45 GHz. The principal advantage of insect control by this electromagnetic method is the absence of any harmful chemical residues. The disadvantage would be an increased cost of grain. However, dropping prices of commercially available RF and microwave power sources and increasing social awareness of chemical intake might make this method attractive in the future. More recent experiments with combined 2.45 GHz mi·)wave heating and partial vacuum indicated some promise in controlling stored-grain i.sects [4], but future application will be dictated by economic factors.

Seed Treatment. Seeds of many legumes often present a germination problem owing to a large number of hard seeds. While these seeds are viable, an impermeable seedcoat prevents the entry of moisture necessary to initiate germination, and consequently the seeds germinate late and plants may not have sufficient time to mature by harvest time. Such seeds, when heated to an appropriate temperature, show greatly improved germination without any other undesirable side effects. Extensive studies were conducted of alfalfa seed germination after treatment at various frequencies [4] discovering favorable response of this species to RF and microwave exposures. However, results obtained for other plant species were less favorable and for some no change was observed. More studies are needed to explain interaction between biological mechanisms and electromagnetic irradiation, possibly even on a molecular and cellular level.

Moisture Determination. Microwave energy can be used in two different ways for moistur·: content determination in agricultural products. The correlation of dielectric properties with moisture content permits the sensing of moisture through electronic measurement of related quantities and this approach will be discussed in more details in the next section. The other means used for moisture determination has been the rapid drying of samples in RF dielectric heaters or microwave ovens to low levels and determining moisture content by sample weight loss and calibration to a standard air or vacuum oven method. An essential advantage of such procedure is its speed when compared with conventional methods of grain drying: several minutes instead of 19 hours (wheat) or 72 hours (corn). Practical application of this method to a given commodity requires careful selection of drying rate and drying time, consideration of other factors pertinent to the commodity, and subsequent calibration of the equipment against an accepted standard method of moisture determination.

Low-Power Applications

Product Quality Assessment. Dielectric properties of some agricultural products might be useful for rapid nondestructive quality evaluation. However, single-frequency measurements used in the past could not reveal correlations that were useful for detecting maturity characteristics in peaches or the "hard core" in sweet potatoes. With the developement of computer-controlled broadband measuring methods, further studies may provide correlations that could be used for these purposes.

Moisture Content and Bulk Density. The moisture content of material, M, expressed in percentage wet basis, is defined as

$$M = \frac{m_w}{m_w + m_d} \times 100 = 100\ \xi \tag{25}$$

that is, as the ratio of the mass of water, m_w, to the total mass of the wet material, with m_d being the mass of dry material and ξ is the fractional moisture content on a wet basis. Considering a unit volume of material, v, Eqn. (25) may be rewritten in the form

$$\xi = \frac{\frac{m_w}{v}}{\frac{m_w}{v} + \frac{m_d}{v}} = \frac{k}{\rho} \tag{26}$$

where k is the water concentration in the unit volume and ρ is the density of the moist material.

Changes of an electrical signal interacting with a moist material, regardless of operating frequency, are proportional to the water concentration, k, and are affected only to a small extent by the mass of dry material, m_d. Thus, when k may be determined from electrical measurement, it is evident from Eqn. (26) that determination of moisture content, ξ, requires the density of the wet material, ρ, to be known. This information can be obtained from a separate density measurement, for example, by weighing a sample of given volume or by using a γ-ray density gauge. Fluctuations in the density of the test material produce effects similar to changes in water content and therefore contribute to a measurement error. This error can be limited or eliminated only if the mass of the wet material in the measuring space is held constant during the calibration procedure and during the measurement. Another way to resolve the density-variation problem is to find a function correlating the material moisture content with measured electrical material parameters, a so-called *density-independent* function. Efforts to eliminate effects of density fluctuations in the moisture content measurements have been made for the last twenty years and some results are presented below.

The dispersion and dissipation of electromagnetic energy interacting with dielectric material depends upon the shape, dimensions and relative permittivity of the material. When the moisture content of the material changes, a change is reflected in the wave parameters. Because the relative permittivity of water differs from that of most hygroscopic dielectric materials, its effect can be separated from the effect of the dry dielectric material. In general, this may be expressed in a functional form as

$$\alpha = \Phi_1(m_w, m_d) \qquad \text{and} \qquad \beta = \Phi_2(m_w, m_d) \tag{27}$$

where α and β are any two descriptive electromagnetic wave parameters. For example, such pairs can include attenuation and phase constants of a plane-wave, or resonant frequency shift and change in the Q-factor of a resonant microwave cavity. Regardless of the complexity of the analytical expressions described by Eqn. (27), it is generally possible to solve the two equations and to express the mass of water and the mass of dry material in terms of two measured parameters in the form

$$m_w = \Psi_1(\alpha, \beta) \qquad \text{and} \qquad m_d = \Psi_2(\alpha, \beta). \tag{28}$$

Substituting the analytical expressions corresponding to Eqns. (28) into (25), the general expression for moisture content of a material can be written as

$$M = \frac{\Psi_1(\alpha, \beta)}{\Psi_1(\alpha, \beta) + \Psi_2(\alpha, \beta)} \times 100 \tag{29}$$

which contains only the wave parameters determined experimentally, and is totally independent of the material density. Moreover, the denominator of Eqn. (29) is the density of the material under test, so it can be determined independent of the moisture

content from the expression:

$$\rho = \frac{m_w}{v} + \frac{m_d}{v} = \Psi_1(\alpha,\beta) + \Psi_2(\alpha,\beta) \tag{30}$$

and then used for correction or other purposes.
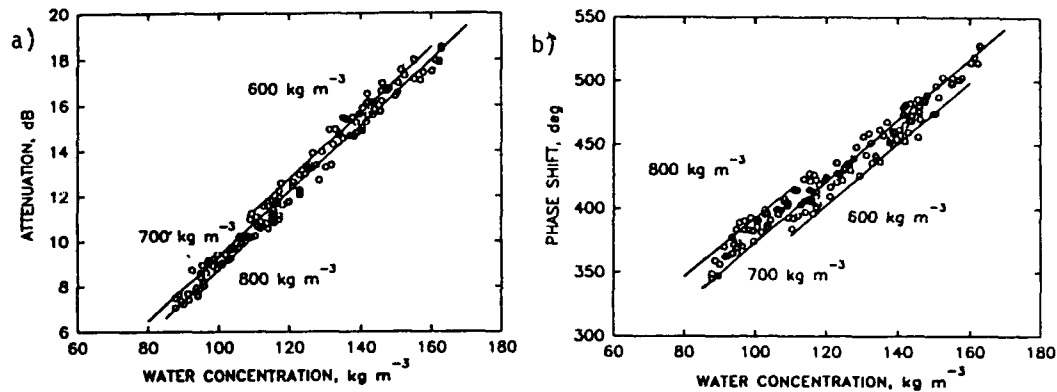
Fig. 4. Attenuation (a) and phase shift (b) of soft red winter wheat as a function of water concentration measured at 4.8 GHz and 24°C on layer of grain 87 mm thick. Lines are for constant values of dry -material density (m_d/v) indicated.

In practice, the equations corresponding to Eqns. (27) are linear over the entire interesting range of moisture content. For example, the results for a layer of wheat irradiated by a plane wave in free space at 4.8 GHz, are shown in Fig. 4. The two linear equations fitting the experimental results with high statistical significance have a form

$$A = -1.17 + 0.1442 \frac{m_w}{v} - 0.0056 \frac{m_d}{v} \qquad r = 0.9928$$

$$\phi = 4.0 + 2.401 \frac{m_w}{v} + 0.1845 \frac{m_d}{v} \qquad r = 0.9896$$

where A is the attenuation in dB, ϕ is the phase shift in degrees, r is the correlation coefficient, and concentration of water and density of dry material are expressed in $kg \cdot m^{-3}$. According to the theoretical considerations, use of Eqns. (29) and (30) provides the expressions for the moisture content and grain bulk density in the form

$$M = \frac{3.733\phi + 123.2A + 129.7}{\phi - 14.8A - 21.362} \quad [\%] \tag{31}$$

$$\rho = 3.74\phi - 55.35A - 79.93 \quad [kg/m^3] . \tag{32}$$

Fig. 5. Distribution of differences: (a) between oven moisture content determination and moisture content calculated for hard red winter wheat from calibration equation (31); (b) between measured grain bulk density and the density predicted from calibration equation (32).

Verification of calibration equations in the form of the above expressions is usually performed with other wheat cultivars harvested the following year. Data from verification measurements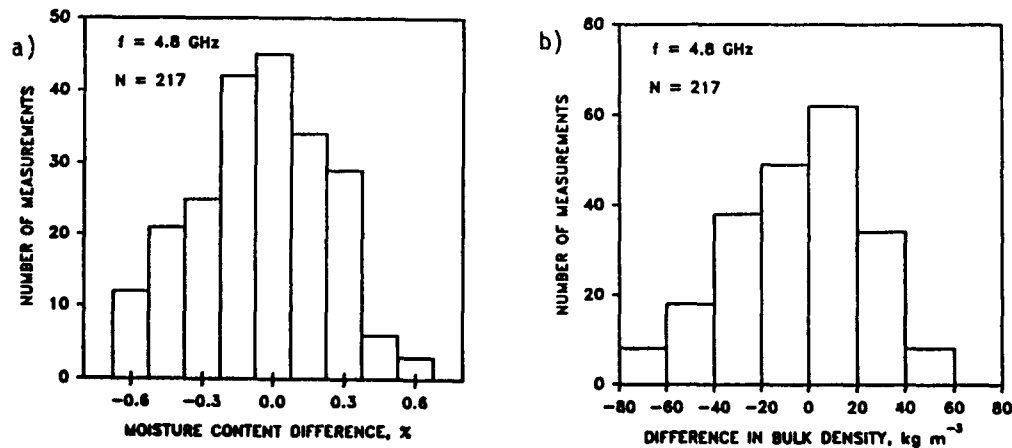 are used in calibration equations, e.g., (31) and (32), and compared with oven moisture and bulk density determinations. The histograms presented in Fig. 5 show the distribution of differences betw. ·;; oven moisture determination and calculated moisture content for these data, as well as the differences between densities determined by weighing and by calculation with Eqn. (32). The mean value of differences between oven moisture and calculated moisture (bias) was −0.032% moisture, and the standard deviation of the differences (standard error of performance, SEP) was 0.271% moisture. The SEP for the bulk density determination was 18.6 kg·m^{-3}, while the bias was −1.6 kg·m^{-3}. .

Moisture Content and Mass of Kernels. Moisture content and mass of single seeds, grain kernels and nuts can be determined using microwave resonators, provided that the measured object is small compared to the volume of the cavity and its loss factor complies with the $\epsilon'^2 >> \epsilon''^2$ requirement. This means that the resonator perturbation theory can be applied and the following expressions hold

$$\Delta F \simeq 2(\epsilon' - 1)\, Kf_o \left(\frac{v_s}{v_o}\right) \qquad (33)$$

$$\Delta T \simeq 4\epsilon''\, K^2\, Q_o \left(\frac{v_s}{v_o}\right) \qquad (34)$$

where $\Delta F = f_o - f_s$ is the shift of resonant frequency, with subscripts o and s refering to the empty cavity and the cavity loaded with a sample (kernel) at the center of the cavity, respectively. Energy dissipated in the kernel is expressed as a change in the cavity Q-factor:
$$\frac{1}{Q_o} - \frac{1}{Q_s} = \frac{1}{Q_o}\left(\frac{V_o}{V_s} - 1\right) = \frac{\Delta T}{Q_o}.$$
Here V denotes the voltage transmission coefficient at resonance, $\Delta T = 10^k - 1$ is the transmission factor, and $k = \frac{1}{20}(S_{21o} - S_{21s})$, with S_{21} being the voltage transmission coefficient at resonance, expressed in decibels.
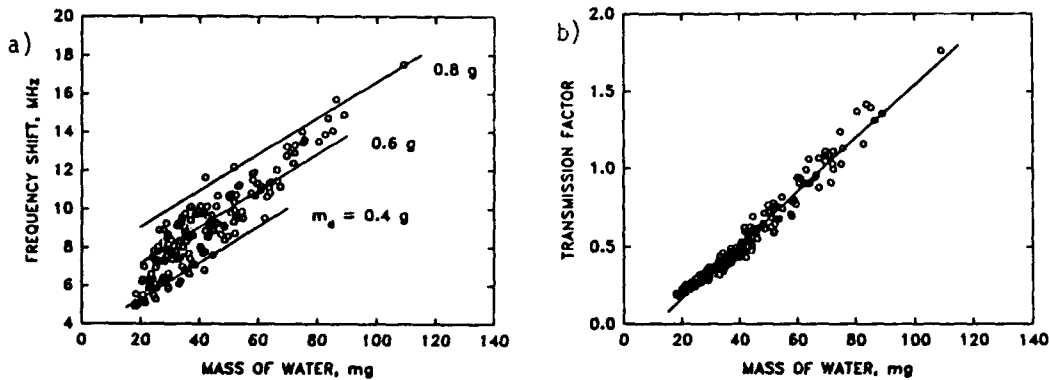
Fig. 6. Dependence of (a) cavity resonant frequency shift and (b) cavity transmission factor on water content in single peanut kernels.

Following the general considerations, experimental results for peanut kernels measured at various moisture levels in the S-band cavity ($f_o = 3175.9$ MHz, $Q_o = 865$) and shown in Fig. 6, can be fitted by the two following linear equations:

$$\Delta F = -0.315 + 0.09456\, m_w + 0.00937\, m_d \qquad r = 0.9930$$

$$\Delta T = -0.057 + 0.01816\, m_w - 0.00028\, m_d \qquad r = 0.9877$$

where m_w and m_d are expressed in milligrams. Then these two equations were solved for m_w and m_d to provide the calibration equations for the wet kernel mass, m_m, and the kernel moisture content, in the form:

$$m_m = 91.365\Delta F - 416.93\Delta T + 16.84 \quad [\text{mg}] \qquad (35)$$

$$M = \frac{0.1635\Delta F + 4.489\Delta T + 0.289}{0.0914\Delta F - 0.4169\Delta T + 0.0168} \quad [\%] \tag{36}$$

Verification of these calibration equations was carried out with peanuts harvested the following year. For a total of 318 data points, the predicted values of peanut moisture co .tent agreed well with those determined by the standard method, with bias of -0.235% moisture and standard error of performance of 0.45% moisture content. It should be noted that the uncertainty in the standard oven moisture determination is generally $\pm 0.2\%$ moisture content. The predicted values of peanut mass agreed with those obtained by weighing with an accuracy of ± 25 mg (two standard deviations) in the whole range of moisture content from 4 to 14%. As before, the distributions of differences between measured and calculated values for single peanut kernel mass and moisture content determination have a normal character, because all experimental errors are of random character.

Similar experiments have been carried out for soybeans, corn and wheat in various resonant cavities appropriate for the kernel sizes. Any speculation on the physical meaning of the numerical coefficients in the calibration equations such as Eqns. (35) and (36) should await additional experiments. In all cases, however, the linear equations were the most appropriate for fitting the experimental data in the range of interesting moisture content. This suggests that in these stored agricultural products water is predominantly in bound form.

Sorting Biological Objects. Perturbation of microwave resonant cavities by agricultural products was utilized to sort potatoes from stones and clods, to grade eggs and apples [2]. Further applications are limited only by imagination and ingenuity of scientists and engineers.

REFERENCES

[1] W. Schilz and B. Schiek, "Microwave systems for industrial measurements", *Advances in Electronics and Electron Physics*, vol. 55, pp. 309-381, 1981.

[2] M.A. Stuchly and S.S. Stuchly, "Industrial, scientific, medical and domestic applications of microwaves", *IEE Proc.*, vol. 130, pt. A, no. 8, pp. 467-503, 1983.

[3] J.F. Alder, M.F. Brennan, I.M. Clegg, P.K.P. Drew and G. Thirup, "The application of microwave-frequency spectrometry, permittivity and loss measurements to chemical analysis", *Trans. Inst. Meas. & Control*, vol. 5, no. 2, pp. 99-111, 1983.

[4] S.O. Nelson, "Potential agricultural applications of RF and microwave energy", *Trans. of the ASAE*, vol. 30, no. 3, pp. 818-831, 1987.

[5] A.W. Kraszewski, "Microwave aquametry – needs and perspectives", *IEEE Trans. on Microwave Theory Techn.*, vol. MTT-39, no. 5, pp. 828-835, 1991.

[6] R.D. Smith, "Present and future uses of microwave power", in *Microwave Processing of Materials II*, W.B. Snyder, W.H. Sutton, M.F. Iskander and D.L. Johnson (eds.), MRS Symp. Proc. vol. 189, pp. 383-390, Pittsburgh, PA: MRS, 1991.

[7] A.W. Kraszewski and S.O. Nelson, "Composite model of the complex permittivity of cereal grain", *J.Agric.Eng. Res.*, vol. 43, pp. 211-219, 1989.

[8] M.C. Dobson and F.T. Ulaby, "Active microwave soil moisture research", *IEEE Trans. on Geosci. Remote Control*, vol. GE-24, no. 1, pp. 23-36, 1986.

[9] F.T. Ulaby and R.P. Jedlicka, "Microwave dielectric properties of plant material", *IEEE Trans. on Geosci. Remote Control*, vol. GE-22, pp. 530-535, 1982.

[10] C.R. Vaughn, "Birds and insects as radar targets: A review", *Proc. IEEE*, vol. 73, no. 2, pp. 205-227, 1985.

[11] S.S. Stuchly, M.S. Sabir and A. Hamid, "Advances in monitoring of velocities and densities of particulates using microwave Doppler effect", *IEEE Trans. on Instrum. Meas.*, vol. IM-26, no. 1, pp. 21-24, 1977.

[12] S.O. Nelson, "Insect-control studies with microwave and other radiofrequency energy", *Bull. Entomol. Soc. Amer.*, vol. 19, no. 3, pp. 157-163, 1973.

[13] A.W. Kraszewski and S.O. Nelson, "Wheat moisture content and bulk density determination by microwave parameters measurement", *Canadian Agric. Eng.*, vol. 34, no. 4, pp. 327-335, 1992.

[14] A.W. Kraszewski and S.O. Nelson, "Nondestructive microwave measurement of moisture content and mass of single peanut kernels", *Trans. ASAE*, vol. 36, no. 1, pp. 127-134, 1993.

WAVE APPROACH TO CAD NOISE ANALYSIS, MODELING AND MEASUREMENT
OF MICROWAVE NETWORKS

Janusz A. Dobrowolski, Senior Member, IEEE

Abstract-The paper presents noise wave approach applied to CAD analysis, design, modeling, and measurement of microwave networks. Methods are presented for noise analysis and noise power sensitivity analysis of microwave circuits with any topology. Algorithms for noise figure and noise parameters computations of two-ports with any internal topology are discussed. The calculation of the noise wave correlation matrices of passive multiports and active two-ports are given. Noise wave parameters for MESFETs and HEMTs based on a small signal equivalent circuit are derived. Finally a noise wave experimental technique for active two-port noise parameter measurements is presented.

I. INTRODUCTION

There are many ways and alternatives for the characterization of noise in linear microwave networks. The most commonly used representations of noise generated in networks are equivalent voltage or current sources, or combinations of both. Impedance and admittance matrix representations, chain matrix representation and few others are used in CAD analysis of noisy networks [1].

At microwave frequencies, however, a wave representation of noise is advantageous. It permits to use scattering matrices and signal flow graphs for noise computations leading to advantages in CAD of microwave networks. In the noise wave representation a circuit element's noise is described using waves that emanate from its ports [3].

As it is shown in Fig.1, noise generated in a two-port element is represented by two mutually correlated noise wave sources c_1 and c_2, one source at each port. The linear matrix equation describing the noisy two-port is

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{1}$$

Warsaw University of Technology, Institute of Electronics Fundamentals, Nowowiejska 15/19, 00-665 Warsaw, Poland.
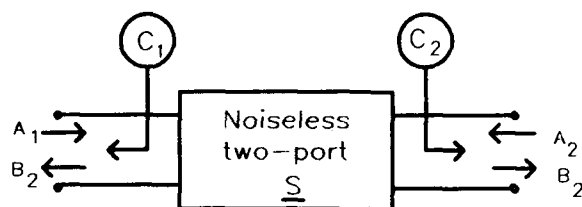
Fig. 1. Noise wave and scattering matrix representation of a two-port.

The ingoing and outgoing noise waves A_i and B_i at the i-th port are defined in terms of port terminal noise voltage V_i and noise current I_i, and a complex reference (normalization) impedance Z_{Ri}

$$A_i = \frac{V_i + Z_{Ri}\, I_i}{2\sqrt{Re\{Z_{Ri}\}}} \qquad B_i = \frac{V_i - Z_{Ri}^*\, I_i}{2\sqrt{Re\{Z_{Ri}\}}} \tag{2}$$

where the asterisk indicates the complex conjugate and Re the real part. Equations (2) are identical to the standard definition of power waves for deterministic signals [2]. The noise waves are time varying complex variables characterized by a correlation matrix C_s given by

$$C_s = \begin{bmatrix} \overline{c_1 c_1^*} & \overline{c_1 c_2^*} \\ \overline{c_2 c_1^*} & \overline{c_2 c_2^*} \end{bmatrix} = \begin{bmatrix} \overline{|c_1|^2} & \overline{c_1 c_2^*} \\ \overline{c_2 c_1^*} & \overline{|c_2|^2} \end{bmatrix} \tag{3}$$

where the overbars indicate time averaging. The diagonal terms of C_s give the noise power deliverable to the matched terminations of ports in a 1-Hz bandwidth. The off diagonal terms are correlation products. The noise wave correlation matrices C_s are Hermitian matrices. Their components are referred to as noise wave parameters.

Scattering parameters as well as noise parameters of a linear nultiport are frequency dependent. Full characterization of signal and noise properties of a multiport requires knowledge of

the scattering matrix and of the noise wave correlation matrix at each frequency of interest. The power of the wave representation of noise lies in its compatibility with the scattering matrix description of microwave networks. This permits noise analysis problems to be formalized and solved using scattering parameters. Microwave devices are most conveniently characterized in terms of their scattering matrix components. Present day availability of accurate scattering parameter measurement methods and systems contributes to the accuracy of the noise analysis. Scattering matrix parameters are numerically stable. The reflections and resonances common in microwave networks do not change dramatically numerical values of these parameters. This fact limits the value range of noise wave quantities and makes them appropriate for CAD applications. The wave interpretation of noise may be also used in alternative noise parameter measurement methods [4],[5],[11].

This presentation discusses applications of the noise wave representation in CAD analysis, modeling, and measurement of microwave devices and networks.

Section II presents two noise wave CAD analysis methods. Noise calculations for interconnected multiport networks with any topology are presented. Algorithms for two-port noise figures of-merit computations are discussed.

Section III presents noise power sensitivity analysis method. Application to the CAD of low-noise circuits with minimum noise figure using efficient optimization methods is envisaged. Noise power sensitivities are applicable also in tolerance analysis and design of microwave networks.

In Section IV noise wave modeling of passive and active circuit elements is discussed. Standard noise parameters F_{min}, Γ_{opt} and R_n received from measurements or equivalent small signal circuits of MESFETs and HEMTs are used to derive noise wave correlation matrices.

Section V describes a very simple and convenient method for direct measurement of noise wave parameters. In its principle the method is similar to the six-port measurement techniques. It does not use expensive source-pull tuners required in conventional techniques to measure standard noise parameters F_{min}, Γ_{opt}, and R_n.

II. NOISE WAVE NETWORK ANALYSIS

The noise analysis of a network is understood as a method for computing noise parameters such as noise figure, four noise parameters of the overall networks etc.

In the analysis, it is assumed that each linear noisy network may be represented as the interconnection of multiport elements [8].
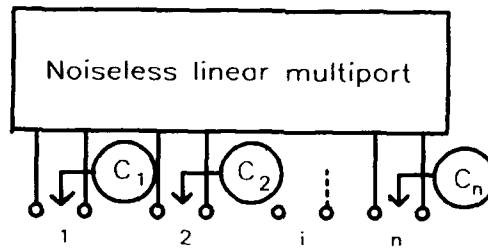
Fig. 2. Multiport noise wave representation.

As it is shown in Fig. 2, noise generated in an element is represented by noise wave sources, one source at each port. In matrix notation, a set of linear equations which relate complex amplitudes of noise waves at ports of an element has the form

$$B^{(k)} = S^{(k)} A^{(k)} + C^{(k)} \tag{4}$$

where $S^{(k)}$ is the scattering matrix of the k-th element, $A^{(k)}$ and $B^{(k)}$ are vectors of incident and reflected noise waves at its ports, and $C^{(k)}$ is a vector of mutually correlated noise wave sources representing noise generated in the element [8].

The multiport element noise wave correlation matrix C_s of each element is given by

$$C_s = \overline{cc^+} \tag{5}$$

where the bar indicates the time averaged correlation product and the dagger the Hermitian conjugate. At a given frequency the signal and noise properties of a linear circuit element are fully described by the scattering matrix S and by the noise wave correlation matrix C_s.

A. CONNECTION SCATTERING MATRIX METHOD

Fig. 3 presents a general networks composed of m elements connected together by their ports. Considering all m elements (multiports) of the network, we have a set of linear equations whose matrix form is

$$B = S A + C \tag{6}$$

where

$$A = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ A^{(k)} \\ \cdot \\ \cdot \\ \cdot \\ A^{(m)} \end{bmatrix} \quad B = \begin{bmatrix} B^{(1)} \\ B^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ B^{(k)} \\ \cdot \\ \cdot \\ \cdot \\ A^{(m)} \end{bmatrix} \quad C = \begin{bmatrix} C^{(1)} \\ C^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ C^{(k)} \\ \cdot \\ \cdot \\ \cdot \\ C^{(m)} \end{bmatrix} \tag{7}$$

and

$$S = \begin{bmatrix} S^{(1)} & 0 & \cdots\cdots\cdots & 0 \\ 0 & S^{(2)} & & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ 0 & & S^{(k)} & 0 \\ \cdot & & & \cdot \cdot \\ \cdot & & & \cdot \\ 0 & \cdots\cdots 0 & \cdots\cdots & S^{(m)} \end{bmatrix} \tag{8}$$

The connections between ports impose constraints on the vectors A and B which can be represented as a matrix equation

$$B = \Gamma A \tag{9}$$

where Γ is the connection matrix of the analyzed network.

It is assumed in the analysis that all ports of connected ports have the same, real reference impedances. In such a case the elements of the connection matrix Γ are all zero except the 1's in the entries corresponding to pairs of adjacent ports.

After eliminating of the vector **B** from (6) and (9) we have

$$W A = C \tag{10}$$

where

$$W = \Gamma - S \tag{11}$$

is the connection scattering matrix of the network.

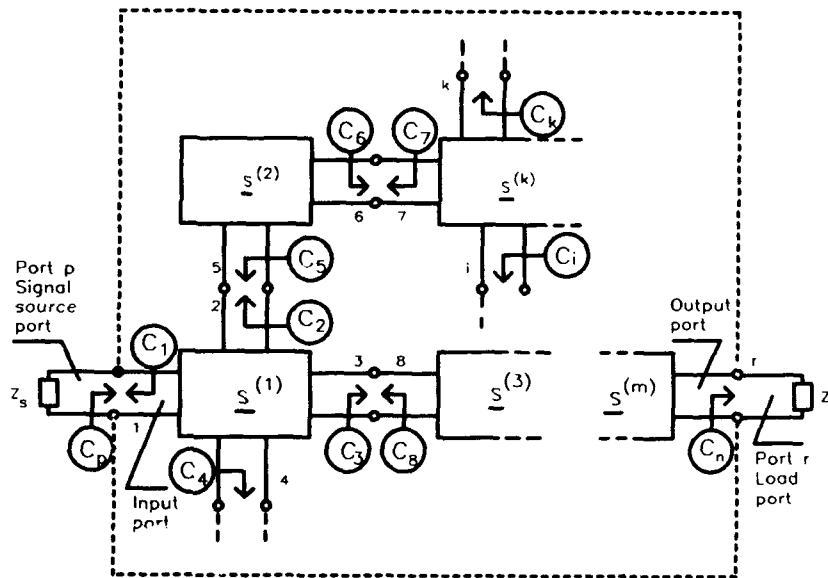Fig. 3. Equivalent circuit of a multiport network with noiseless elements and noise wave sources at each port.

Using (11) we are able to get a correlation matrix of the incident noise waves at all circuit ports. Because

$$A = W^{-1} C \tag{12}$$

then

$$\overline{AA^+} = W^{-1} \overline{CC^+} (W^{-1})^+ = W^{-1} C_s (W^{-1})^+ \tag{13}$$

where the daggers indicate the complex conjugate transpose of the vectors and matrices.

In (13)

$$C_s = \overline{CC^+} = \begin{bmatrix} C_s^{(1)} & 0 \ldots\ldots\ldots 0 \\ 0 & C^{(2)} & & 0 \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot \\ 0 & & & C^{(k)} & 0 \\ \cdot & & \cdot & \cdot \\ 0 \ldots\ldots 0 \ldots C^{(m)} \end{bmatrix}$$ (14)

is the correlation matrix of the noise sources representing noise generated in all circuit elements. Because the noise wave sources $C^{(i)}$ of i-th element are uncorrelated with those of any other circuit element the correlation matrix C^s is a block diagonal matrix with noise correlation matrices of individual network elements at its main diagonal.

To compute the noise wave parameters related to the input and output ports of a general multiport network the reflection coefficient S_{pp} of the input port termination and the reflection coefficient S_{rr} of the output port termination must be set to zero. It must be assumed also that both these terminations are noise free what means that two diagonal elements of matrix C_s relative to ports belonging to the signal generator impedance and to the load impedance are set to zero.

The correlation matrix $\overline{AA^+}$ fully describes noise properties of the analyzed network. In the case of noise analysis of a resulting two-port, four terms of the $\overline{AA^+}$ matrix

$$C_{net} = \begin{bmatrix} (\overline{AA^+})_{pp} & (\overline{AA^+})_{pr} \\ (\overline{AA^+})_{rp} & (\overline{AA^+})_{rr} \end{bmatrix}$$ (15)

corresponding to the input port and to the output port of the network describe noise properties of the resulting two-port. They may be used for computation of noise performance of the network such as noise figure, noise temperature or standard noise parameters of the overall network [7].

B. MULTIPORT CONNECTION METHOD

Computation of the scattering matrix of a large network can be achieved by calculating scattering matrices for each of its elements separately. Subsequently, the elements are interconnected and, finally, the scattering matrix relative to the external ports of the complete circuit is determined.

Fig. 4 illustrates two multiports connected together by internal ports. Evaluation of the scattering matrix of the entire network using scattering matrices of individual multiports is effected by partitioning the scattering matrix T of the embedding network into submatrices in the form [11]

$$\begin{bmatrix} b_e \\ b_i \end{bmatrix} = \begin{bmatrix} T_{ee} & T_{ei} \\ T_{ie} & T_{ii} \end{bmatrix} \begin{bmatrix} a_e \\ a_i \end{bmatrix} \tag{16}$$

where a_i, b_i and a_e, b_e are signal waves at the i internal connections between subnetworks S and T and at the e external ports of the subnetwork T.

The scattering matrix S_{net} referred to the external ports of the network may be computed from the equation

$$S_{net} = T_{ee} + \Lambda\ S\ T_{ie} \tag{17}$$

where

$$\Lambda = T_{ei}\ [I - S\ T_{ii}]^{-1} \tag{18}$$

For a network comprising many components, the computation time is, however, too long due primarily to computation of matrix Λ which requires the inversion of a matrix with order equal to the number of the internal connected ports. Great reduction of the computation time may be obtained by connecting the m component multiports of the complete network two at a time and determining the S_{net} matrix of the resulting subnetwork every time. For a network composed of m component multiports, $(m - 1)$ applications of (17)-(18) computes the overall S_{net} of the complete network.

Noise analysis of a microwave network may be performed also using a principle of the multiport connection method applied for the deterministic analysis. Computation of the noise wave correlation matrix of the entire network related to the input and

output port may be performed the same way, through one or more applications of the multiport connection network problem illustrated in Fig. 4.

It is assumed now that the two subnetworks have known noise wave correlation matrices C^s, and C_t, respectively. The noise wave correlation matrix C_t of the subnetwork T is partitioned in the same way as in the case of scattering matrix S_t in (16). We have

$$C_{net} = \begin{bmatrix} \overline{c_e c_e^+} & \overline{c_e c_i^+} \\ \overline{c_i c_e^+} & \overline{c_i c_i^+} \end{bmatrix} \tag{19}$$

The resulting noise wave correlation matrix is

$$C_{net} = \Lambda \, C_s \, \Lambda^+ + [\, I \mid \Lambda \, S \,] \, C_t \, [\, I \mid \Lambda \, S \,]^+ \tag{20}$$

In (20) I is the identity matrix, and Λ is the matrix given by (18).
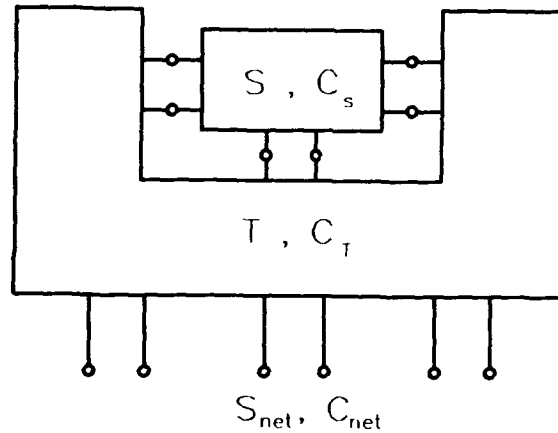
Fig. 4. A subnetwork characterized by the scattering matrix S and the noise wave correlation matrix C_s connected with a subnetwork characterized by the scattering matrix T and the correlation matrix C_t.

As it has been mentioned earlier, the majority of computation time in deterministic analysis is spent on matrix inversion required for matrix Λ, and for the computation of product ΛS. By comparing (17) and (20), we find that matrix Λ and matrix product ΛS are common to both S_{net} and C_{net} expressions. This significantly

reduces computational effort when performing a noise analysis and a deterministic analysis at the same time. The order according to which the network components are connected to each other determines the total number of long arithmetic operations (multiplications and divisions) required to compute the S_{net} and C_{net} matrices for a network with a given topology. The algorithm for near optimal ordering of the connection sequence calls for, at each step, connecting the two components whose resulting multiport has the smallest number of ports [15].

C. TWO-PORT NETWORK ANALYSIS

Consider now resulting two-port network presented in Fig. 5.

Noise performance figures-of merit of this circuit are calculated from the final scattering matrix S_{net} and noise wave correlation matrix C_{net} that result· from a multiport network deterministic and noise analysis. In terms of the scattering and noise wave parameters presented in (17) and (20) excess noise figure $F_e = F - 1$ is given by

$$kT_o\,F_e = \frac{c_{11}\left|\dfrac{s_{21}\,\Gamma_s}{1 - s_{11}\,\Gamma_s}\right|^2 + c_{22} + 2\,Re\left\{c_{12}\,\dfrac{s_{21}\,\Gamma_s}{1 - s_{11}\,\Gamma_s}\right\}}{\left|\dfrac{s_{21}}{1 - s_{11}\,\Gamma_s}\right|^2 (1 - |\Gamma_s|^2)} \quad (21)$$

where Γ_s is the reflection coefficient of the signal generator impedance, T_o is standard temperature and k is Boltzmann's constant. A 1-Hz bandwidth is assumed in equation (21).
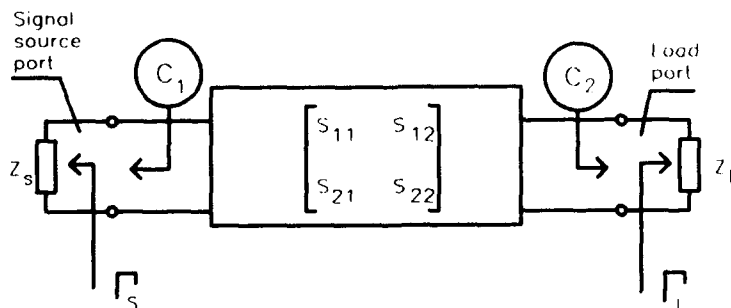

Fig. 5. Resulting two-port network.

Using the C_{net} matrix entries, the standard noise parameters $M_{min} = F_{min} - 1$, Γ_{opt} and R_n may also calculated. Equations for the conversion between these parameter sets are given below:

$$kT_o \, M_{min} = \frac{\overline{|c_2|^2} - \overline{|c_1 s_{21} - c_2 s_{11}|^2} \, |\Gamma_{opt}|^2}{|s_{21}|^2 \, (1 + |\Gamma_{opt}|^2)} \qquad (22)$$

$$\Gamma_{opt} = \frac{\eta}{2} \left(1 - \sqrt{1 - \frac{4}{|\eta|^2}} \right) \qquad (23)$$

$$\frac{4kT_o \, R_n}{Z_R} = \overline{\left| c_1 - c_2 \left(\frac{1 + s_{11}}{s_{21}} \right) \right|^2} \qquad (24)$$

where

$$\eta = \frac{\overline{|c_2|^2} + \overline{|c_1 \, s_{21} - c_2 \, s_{11}|^2}}{\overline{|c_2|^2} \, s_{11} - \overline{c_1 c_2^*} \, s_{21}} \qquad (25)$$

and Z_R is the reference impedance.

III. NOISE POWER SENSITIVITY ANALYSIS

Noise power sensitivities with respect to network parameters are applicable to noise figure minimization of microwave networks using gradient optimization methods. The noise power sensitivities are applicable for accurate and efficient gradient computation of the noise figure. Knowledge of noise power sensitivities allows us to determine these network parameters whose influence on the noise figures-of merit is the most essential. Noise power sensitivities may be also used in computer aided tolerance analysis and tolerance design procedures for noisy microwave networks.

Let us consider again the equivalent circuit of a noisy network presented in Fig. 4.

If r is the number of a load impedance port of the analyzed network, then the noise power dissipated in the load is [9]

$$P_n = \overline{(\mathbf{A}\mathbf{A}^+)}_{rr} \left(1 - |\Gamma_L|^2 \right) \qquad (26)$$

where $\overline{(\mathbf{A}\mathbf{A}^+)}_{rr} = N_r$ is the r-th diagonal element of the

correlation matrix $\overline{AA^+}$, and Γ_L is the reflection coefficient at the output port load Z_L with respect to the reference impedance Z_R of the load impedance port.

The evaluation of N_r can be derived easily. In fact, if only the r-th diagonal element of the correlation matrix $\overline{AA^+}$ has to be determined, then by letting E_r to be a vector whose elements are all zeros except 1 in the position r,

$$E_r = [0,\ldots,0,1,0,\ldots,0]^T \qquad (27)$$

we have

$$N_r = \left(\overline{AA^+} \right)_{rr} = E_r^+ C_s E_r = \left[(W^+)^{-1} E_r \right]^+ C_s (W^+)^{-1} E_r \qquad (28)$$

This relation may be also written in the form

$$N_r = \left(\overline{AA^+} \right)_{rr} = \Lambda^+ C_s \Lambda \qquad (29)$$

where a vector

$$\Lambda = (W^+)^{-1} E_r = (W^{T*})^{-1} E_r \qquad (30)$$

is the solution vector of a system of equations whose coefficient matrix is equal to the complex transpose of the connection scattering matrix W of the analyzed circuit and having E_r as its right-hand side vector.

Suppose that a parameter p in a given noisy network is to be varied without affecting its topology. Parameter variation will affect the noise power delivered to the load of the circuit. Differentiating (28) with respect to p leads to a equation which relates changes of noise power dissipated in the load to changes in the parameter values of the circuit

$$\frac{\partial N_r}{\partial p} = 2 \, Re \left\{ \Lambda^+ \frac{\partial S}{\partial p} W^{-1} C \Lambda \right\} -2 \, \Lambda^+ Re \left\{ \frac{\partial S}{\partial p} S^+ \right\} \Lambda \qquad (31)$$

In order to evaluate the noise power sensitivities, we need to know the partial derivatives of the element scattering matrices, vector Λ given by the solution of the system of equations

$$W^+ \Delta = E_r \qquad (32)$$

and the inverse of the connection scattering matrix W of the analyzed circuit.

These quantities are also used in the sensitivity analysis of the circuit excited by sinusoidal signal (deterministic analysis). The sensitivity of sinusoidal wave a_r at the r-th port is obtained from

$$\frac{\partial a_r}{\partial p} = \alpha^T \frac{\partial S}{\partial p} a \qquad (33)$$

where a is a solution vector of the system of equations

$$W a = c \qquad (34)$$

of the original circuit and α is the solution vector of the system of equations

$$W^T \alpha = E_r \qquad (35)$$

of the adjoined circuit, in which the left hand side vector is the vector E_r given by (27).

Comparing (32) and (35), we find that the solution vectors of both systems of equations satisfy a relation

$$\Lambda = \alpha^* \qquad (36)$$

which means that only one of the vectors needs to be found. This significantly reduces the computational effort when performing noise analysis, noise power sensitivity analysis and deterministic sensitivity analysis of a network at the same time.

The inverse matrix W^1 of the connection scattering matrix W can be computed very effectively using sparse matrix techniques based on LU factorization or on bi-factorization method [1].

Relation (31) defining noise power sensitivity may be used directly for accurate gradient computation of the noise figure of microwave networks. The theory presented is directly applicable to CAD of low noise microwave networks using gradient optimization method.

IV. PASSIVE AND ACTIVE ELEMENT MODELING

Noise properties of passive and active elements are represented by their noise wave correlation matrices. Entries of these matrices may be expressed in terms of scattering parameters and standard noise parameters.

A. PASSIVE MULTIPORTS

Lossy passive multiports generate only thermal noise. The noise wave correlation matrices of these elements are expressed directly by their scattering matrices [6],[10]

$$C_s = kT(I - SS^+) \tag{37}$$

where k is Boltzmann's constant, I - the identity matrix, T is the physical temperature and S the scattering matrix of the multiport.

It should be mentioned here that, in general, passive elements in a microwave network may have different physical temperatures.

B. ACTIVE DEVICES

Noise generated in active two-ports such as FETs and JBTs is represented by their 2X2 correlation matrices. The entries of these matrices are computed from relations

$$\overline{|c_1|^2} = kT_o \, M_{min} \, (\, |s_{11}|^2 - 1 \,) + \frac{kt \, |1 - s_{11} \, \Gamma_{opt}|^2}{|1 + \Gamma_{opt}|^2} \tag{38}$$

$$\overline{|c_2|^2} = |s_{21}|^2 \left(kT_o \, M_{min} + \frac{kt|\Gamma_{opt}|^2}{|1 + \Gamma_{opt}|^2} \right) \tag{39}$$

$$\overline{c_1 c_2^*} = - \frac{kt \, s_{21}^* \, \Gamma_{opt}^*}{|1 + \Gamma_{opt}|^2} + \frac{s_{11}}{s_{21}} \, \overline{|c_2|^2} \tag{40}$$

and

$$kt = \frac{4kT_o \, R_n}{Z_R} \tag{41}$$

where M_{min} - minimum noise measure, Γ_{opt} - optimum reflection

coefficient of the signal generator, R_n - equivalent noise resistance and $T_o = 290$ K.

The standard noise parameters M_{min}, Γ_{opt} and R_n must be obtained through measurements [16]-[17]. The correlation matrices C_s of active devices may be also expressed by any other set of active two-port noise parameters, for example T_{min}, R_n and Z_{opt} or T_{min}, G_n and Y_{opt} [1].

Noise wave correlation matrices of active devices can be derived also theoretically using small-signal equivalent circuits describing their signal and noise properties. In the case of MESFET and HEMT devices it has shown that two resistors in the equivalent circuit possessing frequency independent effective temperatures fully describe their noise properties [12].
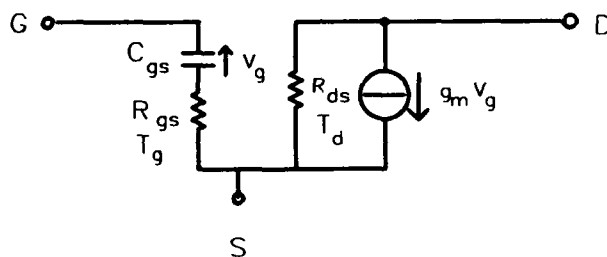
Fig. 6. Intrinsic equivalent circuit for MESFET and HEMT.

Fig. 6 presents intrinsic equivalent circuit of these devices. These two resistors are R_{gs} and R_{ds} with equivalent temperatures, respectively, T_g and T_d. According to this model thermal noise generated by these resistors represent noise properties of field effect devices. In the case of common source configuration entries of noise wave correlation matrix of the device are

$$\overline{|c_1|^2} = kT_g \left(1 - |s_{11}|^2\right) \tag{42}$$

$$\overline{|c_2|^2} = \frac{|s_{21}|^2}{|1 - s_{11}|^2} \overline{|c_1|^2} + kT_d \left(1 - |s_{22}|^2\right) \tag{43}$$

$$\overline{c_2 c_1^*} = \frac{s_{21}}{s_{11} - 1} \overline{|c_1|^2} \tag{44}$$

Correlation between c_1 and c_2 comes from the fact that thermal noise generated by source resistance R_{gs} induces voltage v_i that controls drain current source with transconductance g_m. According to the equivalent circuit scattering parameters of the intrinsic device are

$$S = \begin{bmatrix} \dfrac{1 + j\omega C_{gs}(R_{gs} - Z_R)}{1 + j\omega C_{gs}(R_{gs} + Z_R)} & 0 \\[2ex] \dfrac{-2g_m R_{ds} Z_R}{(R_{ds} + Z_R)[1 + j\omega C_{gs}(R_{gs} + Z_R)]} & \dfrac{R_{ds} - Z_R}{R_{ds} + Z_R} \end{bmatrix} \qquad (45)$$

Practical signal and noise modeling of microwave devices requires consideration of parasitic capacitances, inductances and resistances of the semiconductor chip. Fig. 7 shows realistic equivalent circuit with parasitic elements. Of course, these additional elements will influence signal as well as noise parameters of the model. The resultant scattering and noise correlation matrices of the final equivalent model of a device are computed by applying multiport connection formulas given by (18)-(20). The embedding circuit is usually passive and its noise correlation matrix may be computed from (37). Prior to these computations, the two-port scattering matrix and noise correlation matrix of a three terminal transistor, with one terminal grounded, have to be converted to three-port parameters with the grounded terminal used as a third terminal. The three-port S parameters are found by using the principle that

$$\sum_{i=1}^{N} s_{ij} = \sum_{j=1}^{N} s_{ij} = 1 \qquad (46)$$

for indefinite networks [18].

Similarly, a two-port to three port conversion of the noise wave correlation matrix C_s is achieved by using the principle that is satisfied by the entries of an indefinite N-port noise wave correlation matrix [13]

$$\sum_{i=1}^{N} \overline{c_i c_j^*} = \sum_{j=1}^{N} \overline{c_i c_j^*} = 0 \qquad (47)$$
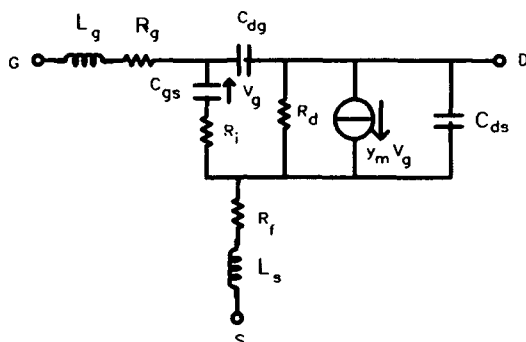
Fig. 7. A more accurate equivalent circuit for the chip.

The noise wave approach simplifies active and passive device modeling by using scattering parameters and physical or equivalent temperature values.

V. NOISE WAVE MEASUREMENTS

The standard two-port noise parameters M_{min}, Γ_{opt}, and R_n, of a microwave transistor are usually determined by measuring noise measure (input noise temperature) for various source reflection coefficients [16],[17]. The procedure requires source pull tuner inserted between the noise source and the device under test. This commonly used technique has a number of disadvantages such as difficulty to automation, repeatability error, restricted ranges of frequency and reflection coefficient tuning and stability problems of measured transistors at low frequencies. The final criticism is that the reflection coefficient and the loss of the tuner have to be frequently calibrated and that the technique is not easily extended to measure noise parameters of multiport networks.

All of these difficulties can be overcome by direct measurement of the noise wave parameters $\overline{|c_1|^2}$, $\overline{|c_2|^2}$ and $c_1 c_2^*$.

Fig. 7 shows a system for measuring the noise parameters of microwave transistors [11].
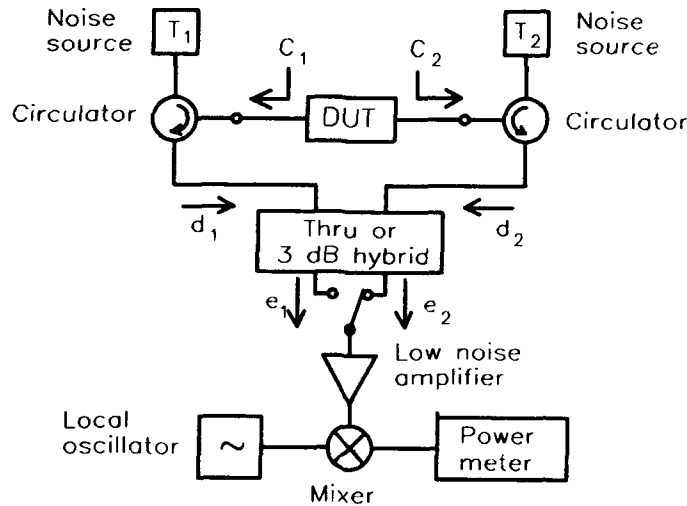
Fig. 7. A system for the measurement of two-port noise
wave parameters.

The noise waves emanating from the device under test are c_1
and c_2. Two circulators located at both ports of the DUT prevent
correlated noise power from a hybrid and an amplifier from
masking the correlated noise power generated by the DUT. The
isolation take place because noise power that originates in the
remainder of the system is absorbed by the matched noise sources
which are connected to the third port of the associated
circulator. Noise waves c_1 and c_2 which emanate from the DUT are
combined with noise waves produced by noise sources and scattered
by the DUT. The resultant noise waves d_2 and d_2 are

$$d_1 = c_1 + \sqrt{kT_1}\, s_{11} + \sqrt{kT_2}\, s_{12} \qquad (48)$$

$$d_2 = c_2 + \sqrt{kT_1}\, s_{21} + \sqrt{kT_2}\, s_{22} \qquad (49)$$

where T_1 and T_2 are effective temperatures of the noise diodes
turned on and off. The measurable quantities of noise waves d_1
and d_2 are

$$\overline{|d_1|^2} = \overline{|c_1|^2} + kT_1\, |s_{11}|^2 + kT_2\, |s_{12}|^2 \qquad (50)$$

$$\overline{|d_2|^2} = \overline{|c_2|^2} + kT_1 |s_{21}|^2 + kT_2 |s_{22}|^2 \qquad (51)$$

$$\overline{d_1 d_2^*} = \overline{c_1 c_2^*} + kT_1 s_{11} s_{21}^* + kT_2 s_{12} s_{22}^* \qquad (52)$$

The noise wave correlation matrix of the DUT can be determined by making measurements using hot and cold values for both T_1 and T_2 and three different circuits ("thru" circuit, 0° / 180° 3 dB hybrid and 0^0 / 90° 3 dB hybrid) inserted between the circulators and the power meter.

a) *With the "thru" circuit:*

$$e_1 = d_1 \qquad\qquad e_2 = d_2 \qquad (53)$$

Four noise power measurements of e_1 and e_2 for hot and cold values of T_1 and T_2 are sufficient to solve (50) and (51) for six unknowns $\overline{|c_1|^2}$, $\overline{|c_2|^2}$, $|s_{11}|^2$, $|s_{12}|^2$, $|s_{21}|^2$, and $|s_{22}|^2$.

b) *With a 0^0 / 180^0 3 dB hybrid:*

$$e_1 = \frac{1}{\sqrt{2}} (d_1 + d_2) \qquad e_2 = \frac{1}{\sqrt{2}} (d_1 - d_2) \qquad (54)$$

and

$$\overline{|e_1|^2} - \overline{|e_2|^2} = 2 \, Re \left\{ \overline{d_1 d_2^*} \right\} \qquad (55)$$
$$= 2 \left[Re \left\{ \overline{c_1 c_2^*} \right\} + kT_1 \, Re \left\{ s_{11} s_{21}^* \right\} + kT_2 \, Re \left\{ s_{12} s_{22}^* \right\} \right]$$

Four noise power measurements of e_1 and e_2 for hot and cold values of T_1 and T_2 allow to solve (55) for unknowns $Re \left\{ \overline{c_1 c_2^*} \right\}$, $Re \left\{ \overline{s_{11} s_{21}^*} \right\}$, and $Re \left\{ \overline{s_{12} s_{22}^*} \right\}$.

c) With a 0_o / 90_o 3 dB hybrid:

$$e_1 = \frac{1}{\sqrt{2}}\,(d_1 + jd_2) \qquad e_2 = \frac{1}{\sqrt{2}}\,(jd_1 + d_2) \qquad (56)$$

and

$$\overline{|e_1|^2} - \overline{|e_2|^2} = 2\,Im\left\{\,\overline{d_1 d_2^*}\,\right\} \qquad (57)$$

$$= 2\left[\,Im\left\{\,\overline{c_1 c_2^*}\,\right\} + kT_1\,Im\left\{\,\overline{s_{11}s_{21}^*}\right\} + kT_2\,Im\left\{\overline{s_{12}s_{22}^*}\right\}\,\right]$$

Four additional noise power measurements permit solution of (57) for unknowns $Im\left\{\,\overline{c_1 c_2^*}\,\right\}$, $Im\left\{\,\overline{s_{11}s_{21}^*}\,\right\}$, and $Im\left\{\,\overline{s_{12}s_{22}^*}\,\right\}$.

In each case of measurements the system of equations is over-determined and statistical analysis may be used in solution procedures. The system can be calibrated by replacing the DUT with a resistive matched attenuator. As it is seen from (37) the noise waves at its ports are correlated with zero phase.

The noise wave method is particularly suitable for the on-chip measurement of transistor noise parameters, the characterization of devices at RF through millimeter wave frequencies and at cryogenic temperatures. Is possible to apply many other configurations for the measurement systems among which the six-port networks are the most interesting [14].

VI. CONCLUSION

The wave representation of noise is a powerful way of alternative analysis, modeling and measurement techniques of microwave or millimeter wave networks. Noise analysis of microwave networks may be performed solely in terms of scattering matrix parameters. Noise wave parameters of passive devices are expressed by scattering parameters only. The expressions describing noise wave parameters for active devices such as microwave FETs and HEMTs are easily derived and interpreted physically. CAD noise analysis and noise power sensitivity analysis of microwave networks with any topology may be performed very effective computationally in parallel with deterministic analysis. The noise parameter measurement techniques based on wave representation of noise may be used at millimeter-wave frequencies. It is especially attractive for on-chip and on-wafer

measurements, measurements at cryogenic temperatures and in cases when the device is relatively inaccessible.

REFERENCES

[1] J.A. Dobrowolski, "Introduction to computer methods for microwave circuit analysis and design", Artech House, Boston - London 1991.
[2] K. Kurokawa, "Power waves and the scattering matrix", IEEE Trans. Microwave Theory Tech., vol. MTT-13, pp. 194-202, March 1965.
[3] P. Penfield, "Wave representation of amplifier noise," IRE Trans. Circuit Theory, vol. CT-9, pp. 84-86, Mar. 1962.
[4] R.P. Meys, "A wave approach to the noise properties of linear microwave devices,"IEEE Trans: Microwave Theory and Tech., vol. MTT-26; pp: 34-37, Jan.1978.
[5] S. Withington, "Scattered noise waves in microwave and mm-wave networks," Microwave Journal, vol. 32, pp.169-178, June 1989.
[6] H. Bosma, "On the theory of linear noisy systems," Philips Res. Repts. Suppl., no. 10, 1967.
[7] H.A. Haus and R.B. Adler, "Optimum noise performance of linear amplifiers," Proc. IRE, vol. 46, pp.1517-1533, Aug. 1958.
[8] J.A. Dobrowolski, "A CAD-oriented method for noise figure computation of two-ports with any internal topology," IEEE Trans. Microwave Theory Tech., vol. MTT-3 i, pp.15-20, Jan. 1989.
[9] J.A: Dobrowolski, "Noise power sensitivities and noise figure minimization of two-ports with any internal topology," IEEE Trans. Microwave Theory Tech., vol. MTT-39, pp.136-140, Jan.1991.
[10] S.W. Wedge and D.B. Rutledge, "Noise waves and passive linear multiports," IEEE Microwave Guided Wave Lett., vol. MGL-1, pp.117-119, May 1991.
[11] S. W. Wedge, D.B. Rutledge, "Wave techniques for noise modeling and measurement", IEEE Trans. Microwave Theory Tech., vol. MTT-40, pp.2004-1012, Nov. 1992.
[12] M.W. Pospieszalski, "Modeling of noise parameters of MESFET's and MODFET's and their frequency and temperature dependence", IEEE Trans. Microwave Theory Tech., vol. MTT-37, pp.1340-1350, Sept. 1989.
[13] T.O. Grosch, L.A. Carpenter, "Two-port to three-port noise wave transformation for CAD applications", IEEE Trans. Microwave Theory Tech., vol. MTT-41, pp. 1543-1548, Sept. 1994.
[14] G.F. Engen, "An improved circuit for implementing the six-port technique of microwave measurements," IEEE Trans. Microwave Theory Tech., vol. MTT-25, pp.1080-1083, Dec. 1977.
[15] V. Monaco, P. Tiberio, "Computer aided analysis of microwave circuits", IEEE Trans. Microwave Theory Tech., vol. MTT-22, pp.249-163, March 1974.
[16] W. Wiatr, "A method of estimating noise parameters of linear microwave two-ports", Ph.D dissertation, Warsaw University of Technology, Warsaw, Poland, 1980 (in Polish).
[17] M. Pospieszalski, "On the measurement of noise parameters of microwave two-ports", IEEE Trans Microwave Theory Tech., vol. MTT_34, pp. 456-458, Apr. 1986.
[18] G.E. Bodway, "Circuit design and characterization of transistors by means of three-port scattering parameters", Microwave J., pp. 55-63, May 1968.

Advances in Technology of Microwave
Submicrometer Devices and Integrated Circuits

Andrzej Jeleński, Lech Dobrzański

Institute of Electronic Materials Technology, 01-919 Warsaw, Wólczyńska 133

Introduction

The purpose of this paper is a brief discussion of recent developments of technology from micro to nanoelectronics. First the HEMT's technology and its application in MMICs will be reviewed based on reports, that appeared in 1993 and 1994. Analysis of these reports will provide informations on present trends and the future progress of the modern microwave electronics.

GaAs, InP and ternary compounds

Almost all modern transistors have channel made of $Ga_{1-x}In_xAs$. The higher is the indium contents in a system the lower is the effective mass of electrons. There are two basic vertical HEMT's structures. The first one is grown on a GaAs semi-insulating wafer, the second on the SI InP substrate. In the former case the heterojunction responsible for separation of dopant and carriers is formed on interface of $GaAlAs/Ga_{1-x}In_xAs$ layers. Typical contents of indium in such a system is 20%. The lattice constant of channel layer is different than in GaAs substrate. There is a stress in a channel material. Relaxation of this stress is possible for indium contents greater than 20%. However, there are reports on so called metamorphic HEMT. GaAs substrate is isolated from channel with thick compositionally graded buffer layer. Indium contents in a channel was enhanced to 29% (ref. 9 in Tab.I). It is the limit of systems grown on GaAs. This systems emerged first, for GaAs technology was mature in comparison to the new one founded on InP.

It was likely, that systems grown on InP substrate can outperform HEMTs made with the old technology. Indium contents can vary in a wide range in a case of a new technology. In particular for 53% of indium a perfect match of lattice constants in substrate and in channel can be obtained.

Reports mentioned in Tab.I with one exception (ref. no 2) describe HEMTs with no kink effect. This effect was a serious drawback of InAlAs/InGaAs/InP system, for drain - to source breakdown voltage of transistor was limited to max. 3V. It seems likely, that "kink effect" do not appear for DMD (double modulation doped) transistor. Some researchers claim to have no kink effect even in SMD (single modulation doped) transistor.

The second serious drawback of InP - based HEMT - gate leak current was reduced by application of new selectively working wet etching solutions. An action of these solutions prevents contact of channel layer with gate electrode at the mesa sidewalls. Furthermore, the quality of the InAlAs layer contacting the gate was significantly improved.

In this way large signal operation of InP - based HEMT is enabled. A year ago this area of application was covered by GaAs - based HEMTs. InP - based HEMTs were believed to suit for low

noise, small signal systems.

Strained or unstrained layers

An unstrained channel should exhibit better noise properties than strained one. In ref.4 Tab.I the direct comparison of k parameter from famous Fukui formula is presented for HEMTs with different indium contents in a channel layer. The lowest k value was obtained for a matched (53% of In) transistor. Researchers working for Toshiba (ref. 2, Tab. I) obtained NF of 0.3 dB at 12 GHz for the matched transistor. Researchers from Avantek - H.P. report NF of 0.23 dB at 12 GHz (ref. 10, Tab. I) for so called pseudomorphic transistor (with strained channel). Parameters mentioned here show a bright prospect of application of InP - based HEMTs in DBS systems.

MOCVD versus MBE technology

MBE technology is superior over MOCVD, but more expensive. The best transistor mentioned in Tab.I was made by MBE in NEC (f max. of 345GHz). The serious effort is made to improve MOCVD technology. Impressive results with MOCVD were obtained at Univ. of Illinois (researchers at present with IBM). The InP - based HEMT exhibits (ref. 8, Tab. I) f_T of 187 GHz. Similar performance was obtained at Wright Patterson Air Base (USA). MOCVD grown HEMT has the highest, ever reported for this technology, alues of f_T and f_{max}: 200 and 230 GHz respectively. In a previous paragraph a MOCVD grown MEMT for low noise application was mentioned (NF of 0.3dB at 12 GHz). MOCVD technology appears to be mature for both professional applications and commercial production.

Metallization for HEMTs

All the transistors mentioned in Tab.I were fabricated using the same old fashioned AuGeNi ohmic metallization. It seems not likely, that this system is used in a serious production process. All the transistors mentioned in Tab.I have gate metallization of TiPtAu. The shape of a gate electrode was in each case mentioned, patterned in PMMA layer with e - beam. Gate metal was than evaporated and lifted off. The gate was recessed by means of the wet etching. Transistors were isolated by wet etching of mesas as well. The gate length shorter than $0.2\mu m$ became a standard. If reported data are reliable, the astonishing uniformity of the HEMT technology emerged recently.

MMICs

Some reports on MMICs fabricated with HEMT process are mentioned in table II. Ka - band is covered with both GaAs and InP based HEMTs. Newest reports claim, that operation at 60-70GHz is now possible with ICs founded on InP system.

Tab. I

Ref.	Technology	Structure	Doping	Kink effect	Lg μm	f_r GHz	f_{max} GHz	Mobility cm²/Vs	NF dB	Author's affiliation
1.	MBE	In AlAs/Ga$_{0.35}$In$_{0.65}$As/InP	DMD	NO	0.2	80	160	7500	0.3 (12GHz)	NTT
2.	MOCVD	In AlAs/GaInAs/INP matched	SMD	YES	0.13		150	10000	-	Toshiba
3.	MOCVD	In AlAs/GaInAs/InP matched	-	NO	0.15	200	230	11900	-	Air Force Wright Lab.
4.	MBE	.../GaIn$_{0.53}$As/InP	-	-	1	26	55	10500	2 (Fukui's k)	Air Force Wright Lab.
		.../GaIn$_{0.82}$As/InP	-	-	1	36	84	12700	3.1 (k)	
		.../GaIn$_{0.7}$As/InP	-	-	1	47	61	13300	3.6 (k)	
		.../GaIn$_{0.8}$As/InP	-	-	1	63	52	13500	5 (k)	
5.	MOCVD	InP/GaIn$_{0.75}$As/InP	DMD	NO	0.5	57	135	9750		Inst.Halb.Aachen
6.	MBE	InAlAs/GaInAs/InP matched	DMD	NO	0.15		345			NEC
7.	MBE	AlGaAs/In$_{0.19}$Ga$_{0.81}$As/GaAs	-	-	0.2	20	100			Univ.Lille
8.	MOCVD	InAlAs/GaInAs/InP matched	SMD	NO	0.15	187		9077		Univ.Illinois-IBM
9.	MBE	In$_{0.29}$Al$_{0.71}$As/In$_{0.3}$Ga$_{0.7}$As/GaAs	SMD	NO	0.4	45	115	7600		Univ.Lille
10.	MBE	InAlAs/InGaAs/InP pseudomorph.	-	-	0.15	185			0.23 (12GHz)	Avantek-H.P.

Tab. II

Ref.	Function	Technology	Lg μm	Band	MAG dB	Author's affiliation
11.	Ampl. 3 stg.	AlGaAs/In$_{0.2}$Ga$_{0.8}$As/GaAs	-	Ka	18	Texas Inst.
12.	Ampl. 3 stg.	InAlAs/GaInAs/InP matched	0.25	26-32GHz	27	Daimler-Benz
13.	Divider	AlGaAs/InGaAs/GaAs	0.15	28-51GHz	-	Fraunhofer Inst.
14.	Ampl. 3 stg.	.../GaInAs/InP	0.1	0.1-75GHz	17	Hewlett-Packard
15.	Ampl.	InAlAs/GaInAs/InP	0.15	60GHz	9	NTT

Lithography

The drive towards higher frequencies and ultra-low scale integration (ULSI) pushed optical lithography well in submicrometer regime. Optical steppers combining high optical aperture with short exposure wavelength (eg.deep. UV at 248 mm) allow the printing of lines 0.3μm thick with classical resists. With liquid phase silylating resists or phase shifting masks even higher resolutions were achieved [16] but will not exceed 0.1μm. Focused ion beams have not yet found wide applications and electron beam direct writing is commonly used. With organic resists such as PPMA the minimal achieved linewidth was around 100Å [17]. To extend this limit researchers tried more resistant inorganic materials, and recently SiO$_2$ exposed through a sacrificial layer to avoid organic contaminations on SiO$_2$ enabled to obtain lines about 10nm in width and spacing on silicon. Other dielectrics can be used for GaAs substrate.

The ultimate - atomic scale resolution can be achieved by scanning tunnelling microscopy, and some experiments show that single atom transport is possible [18]. However this procedure is extremely slow, and since presently most operating nanodevices have still horizontal dimensions larger than 10nm, this technique was not yet utilized.

Device patterning

The most popular is the lift off technique since it is scalable below 100nm. More elaborate techniques such as multilevel resist schemes or dielectric assisted lift-off are needed to fabricate T-shaped gate profiles required to decrease the gate access resistance. Such gates with ~ 50nm footprint have been fabricated. The wet etching, the most popular as can be seen from the above is limited by poor controllability and sloped edges. Therefore in nanotechnology the reactive ion etching (RIE) enabling a much better resolution, vertical anisotropy and high degree of control is becoming increasingly popular in fabrication of nanodevices, despite the radiation damage it produces, and which is actually the main concern. For GaAs and AlGaAs the mixture of methene and hydrogen has shown some advantages over chlorine and fluorine based plasmas commonly used. However an additional thermal annealing is needed to activate the surface layer, deactivated due to the formation of hydrogen - dopant complexes. RIE or H$_2$ plasmas are also used for the electrical isolation of the devices.

Ohmic contacts

Although as can be seen from the above, AuGeNi contacts are still universally used, the nonuniformities of the alloyed region, poor surface morphology and various intermetallic phases limit their scaling down. Some other metals were investigated, and the PdGe contact to GaAs seems to be promising [19].

Where we are?

Substrate materials GaAs and InP are fabricated in the Institute of Electronic Materials Technology. Various epitaxial layers of ternary and quaternary compounds on these materials ca be deposited in this Institute using one of the most performing MOCVD machines in the world: the Aixtron 200. The modern MBE machines in the Institute of Electron Technology and in the Physics Institute of the Polish Academy of Sciences enable the epitaxy on III/V ana II/IV semiconductors.

The 300 nm UV optical lithography enabling fabrication of devices with gate lengths down to $0.5\mu m$, and electron beam writer enabling linewidths down to $0.2\mu m$ are also available in ITME.

Modern patterning wet and dry systems (RIE) are available in ITME and ITE. Examples of a mature $0.5\mu m$ gate technology, and bridges over the GaAs surface are the examples of these possibilities.

References

1. F. Gueissaz *at all*, *Electronics Lett.*, p. 2222 (1993).

2. S. Fujita *at all*, *Electronics Lett.*, p. 1557 (1993).

3. K.B. Chough *at all*, *Electronics Lett.*, p. 1361 (1993).

4. K.B. Chough *at all*, *Electronics Lett.*, p. 1338 (1993).

5. A. Mesquida Kusters *at all*, *Electronics Lett.*, p. 841 (1993).

6. N. Iwata *at all*, *Electronics Lett.*, p. 628 (1993).

7. H. Fawaz *at all*, *Electronics Lett.*, p. 615 (1993).

8. K. Nummila *at all*, *Electronics Lett.*, p. 274 (1993).

9. P. Win *at all*, *Electronics Lett.*, p. 169 (1993).

10. T. Hwang *at all*, *Electronics Lett.*, p. 10 (1993).

11. H.Q. Tserng *at all*, *Electronics Lett.*, p. 304 (1993).

12. J. Dickmann *at all*, *Electronics Lett.*, p. 493 (1993).

13. A. Thiede *at all*, *Electronics Lett.*, p. 933 (1993).

14. C. Madden *at all*, *IEEE Int. Solid-State Conference* (1994).

15. T. Shibata *at all*, *IEEE Int. Solid-State Conference* (1994).

16. K.H. Baik *at all*, *Proc. of SPIE 92*, 362 (1992).

17. A.N. Broers, *J.Electrochem. Soc.*, **128**, 166 (1992).

18. P. Moriarty *at all*, *Appl. Phys. Lett.*, **60**, 2338 (1992).

19. E.D. Marshall *at all*, *Appl. Phys. Lett.*, **47**, 298 (1985).

RECENT ADVANCES IN POWER AMPLIFIER DESIGN METHODOLOGIES

Franco Giannini* and Ernesto Limiti*

Abstract

Next years should see an increasing usage of GaAs-based integrated subsystems, due to the growing importance of mobile communications and related fields. Among such subsytems, high efficiency power amplifier modules will play a real strategic role, being the key component to achieve low cost, low weight, small size, low power consumption and high performances. The design methodologies involved are therefore crucial and cannot be left to the designers' personal experience or skillness. In this paper the relevant existing methodologies for power amplifier design will be commented including a novel one, developed by the authors.

Established techniques

Since 1980 Sechi [1] presented an optimal design technique for high-efficiency quasi-linear (e.g. class-A) power amplifiers achieved in graphical form, with the active device characterized by load impedance contours for constant output power and intermodulation distortion. A design trade-off is performed through the evaluation of the common gradient line, leading to the identification of the optimum load impedance for a fixed value of IMD. The results are however valid for class-A operation of the power stage, at moderate compression levels, and do not account for the possible harmonic loading of the active device.

Another typical class-A approach was introduced by Cripps [2] and successively modified by Kondoh [3]: by means of a simplified device model, ouptut power load-pull contours are generated from simple mathematical expressions; optimum output loading can therefore be determined. However, the validity of the Cripps' approach is restricted to linear operation and, once again, the influence of the harmonic terminations is completely ignored.

Harmonic control techniques are indeed fundamental tools for power amplifiers' performance optimization: since 1967 Snider [4] demonstrated their effectiveness on the maximization of output power and efficiency. In particular, when the optimum efficiency case is concerned, assuming perfect short- and open-circuit terminations at even and odd harmonics respectively, his analysis led to the well-known result of a square waveform as the optimum shape for the output voltage. Anyway, several considerations limit the practical validity of this design approach: firstly, the output conductance of actual devices affects the odd-harmonics voltage components amplitudes, not allowing an unconstrained waveshaping. Secondly, in most practical cases, the complexity of the resulting harmonic-controlling circuitry the difficult control of its impedance at very high frequencies limit the number of controllable harmonics to the first three ones.

More recently, approaches incorporating a more realistic piecewise-linearized device model and accounting for the physical limitations (e.g. breakdown, current saturation, etc.) of the active device have been presented [5-8].

---

* Department of Electronic Engineering, University of Roma "Tor Vergata",
 Via della Ricerca Scientifica, 00133 Roma - Italy

Closed-form expressions for the main output quantities (output power, large-signal gain and power-added efficiency) are obtained as functions of the output current conduction angle, drive level and loading, leading in all cases to design charts indicating the possible trade-offs for specification fulfilment. Moreover, simplified design methodologies overcoming the abovementioned limitations of the Snider's work have been suggested [9,10]. Anyway, all of the abovementioned approaches cannot be used for the final design of a power stage, but only for the investigation of the power capabilities of a device or for preliminary design choices: they are all based on an extremely simplified device model, and the closed-form expressions obtained are just an approximation (more or less accurate depending on the adopted strategy) of the actual behavior of the power device. These *simplified-quasi-nonlinear* methods are therefore very useful to get the necessary physical insight but their inherent inaccuracy discourages their usage for the final design iteration.

Recent procedures

More accurate design procedures have been recently introduced, favoured by the improvements of general-purpose CAD programs [11-13] on one hand, and on the availability of sophisticated load-pull setups [14-16]. All of these procedures can take advantage of exact non-linear analysis or experimental power characterization and are therefore "*full non-linear*" ones.

The most popular non-linear algorithms have been developed in a mixed time-frequency domain (Harmonic balance techniques, [17]), and require an accurate experimental modeling of the active component, in order to extract a large-signal circuit model. The design is performed through repeated analyses until optimum driving and loading conditions are obtained. Anyway, large and expensive equipments are necessary for the model extraction, and considerable computational effort and skilled personnel must be dedicated to the optimization.

The load/source-pull technique is comparatively more straightforward, requires less skilled personnel and no preliminary modeling of the active device. On the other hand, even if posssible in principle, it is difficult to control the harmonic terminations (especially at high frequency of operation) and to explore the low-impedance end for the output load. A typical automated load-pull setup is shown in fig.1. Automated tuners and active load-pulling techniques [18] are now becoming familiar to the microwave community as the key components to increase the speed and flexibility of power characterization via the load-pull method. Nevertheless, the cost of such systems is actually very high and up to now the possibility of harmonic control is still under way.

Both techniques require a high computational or experimental effort, when compared to the dedicated simplified quasi-nonlinear methods. A methodology combining the accuracy of the full non-linear techniques and the speed of the simplified algorithms is therefore highly desirable. In the following, a mixed analysis/optimization algorithm using a non-linear model of the active device will be presented, dedicated to the design of microwave power amplifiers. This algorithm is the core of a design procedure involving the choice of operating class, drive level and optimum loading of a power stage.

Basic assumptions

Two hypotheses are made on the non-linear power amplifier:

a)   The load seen by the non-linear voltage controlled current source (fig.2) must be resistive to ensure high-power and high-efficiency operations. In fact, maximum active power is achieved when output voltage and current are in-phase and their swings maximized. This condition clearly corresponds to have no hysteresis in the load curve (fig.3).

b)   The non-linear reactive elements of the device (i.e. gate-source and gate-drain capacitances) can be replaced by bias-dependent linear equivalent elements when maximum power and efficiency evaluation is concerned. This hypothesis in fact is not, in the general case, required for the algorithm implementation and has been introduced only for the sake of problem complexity reduction. Reactances do not contribute to power generation or dissipation but only to waveform distortion and reactive power handling: in a well-designed amplifier both effects are minimized and compensated by suitable matching at fundamental and harmonic frequencies. The only non-linear element in the equivalent circuit model is therefore the voltage-controlled current source, and any general dependance of the drain current on gate-source and drain-souce voltages can be used, including non-quasi-static dispersion effects [19-21].

Under these hypotheses, a partial matching condition is thus imposed at the output of the device; a second condition of complete transfer of the power available at the input is imposed, implying a conjugate matching at the input port; the matching must be imposed under large-signal operation, simulating the source-pull procedure.

Matching at harmonic frequencies must contribute to produce an hysteresis-free load curve, i.e. a resistive load must be seen by the non-linear current source at every harmonic. A free design parameter is therefore available at each harmonic frequency.

The nonlinear solving system

For a fixed input power, bias point and resistive part of the load, a standard harmonic balance analysis of the circuit requires the solution of a non-linear system of equations. The fulfilment of the conditions described in the previous paragraph requires the solution of additional equations for large-signal circuit match; at the same time however it also reduces the number of harmonic-balancing non-linear equations. In fact, at fundamental frequency the (scalar) equation, imposing the reactive part of the load $X_{gen}^{(1)}$ seen by the controlled current source to be zero, forces the large-signal voltage and current to be in-phase; therefore Kirchhoff's (harmonic-balancing) equation giving the large-signal amplitudes of voltage $V_{gen}^{(1)}$ and current $I_{gen}^{(1)}$ at the same port has now only a real part. The unknowns in the two equations are the reactive matching element in the output network $X_{Load}^{(1)}$, the matching elements in the input network Z_A and the amplitude of the internal drain-source voltage $V_{gen}^{(1)}$:

(1) $$X_{gen}^{(1)} = X_{gen}^{(1)}(X_{Load}, Z_A, g_{m,LS}) = 0$$

(2) $$V_{gen}^{(1)} - R_{gen}^{(1)} \cdot I_{gen}^{(1)} = V_{gen}^{(1)} - R_{gen}^{(1)}(X_{Load}, Z_A, g_{m,LS}) \cdot I_{gen}^{(1)}(X_{Load}, Z_A, V_{gen}) = 0$$

The intrinsic drain current harmonic components $I_{gen}^{(n)}$ (fig.2) are computed from the drain-source and gate-source voltage through a time-domain integration and Fourier transform as in the standard harmonic-balance procedure. The two equations are coupled through the large-signal transconductance, defined as:

$$g_{m,LS} = \frac{I_{gen}^{(1)}(X_L, Z_A, V_{gen})}{V_{gs}(X_L, Z_A, V_{gen})}$$

The additional complex equation imposing complex conjugate match at the input has to be added:

(3) $$Z_A = Z_{in}^*(X_{load}, g_{m,LS})$$

Among the four resulting scalar equations however only (2) requires the costly time-domain integration and Fourier transform of the standard harmonic-balance procedure; the other three are simple frequency-domain matching equations incorporating the large-signal transconductance $g_{m,LS}$.

One 'matching' and one 'harmonic balancing' Kirchhoff's equation at the *internal drain-source port* must similarly be added for each harmonic frequency. The additional unknowns are a (frequency selective) reactive matching element in the output network and the real amplitude of the drain-source voltage harmonic component for each frequency. The complete system is therefore:

$$DC \quad V_{gen}^{(0)} - R_{gen}^{(0)} \cdot I_{gen}^{(0)} = 0 \qquad \text{(scalar)} \qquad \text{(bias conditions + rectification)}$$

$$f_0 \quad \begin{aligned} Z_A &= Z_{in}^* & \text{(complex)} & \quad \text{(large-signal input matching)} \\ X_{gen}^{(1)} &= 0 & \text{(scalar)} & \quad \text{(matching at the current source)} \end{aligned}$$

(4) $$V_{gen}^{(1)} - R_{gen}^{(1)} \cdot I_{gen}^{(1)} = 0 \qquad \text{(scalar)} \qquad \text{(harmonic balancing)}$$

$$\cdot$$
$$\cdot$$

$$n \cdot f_0 \quad \begin{aligned} X_{gen}^{(n)} &= 0 & \text{(scalar)} & \quad \text{(matching at the current source)} \\ V_{gen}^{(n)} - R_{gen}^{(n)} \cdot I_{gen}^{(n)} &= 0 & \text{(scalar)} & \quad \text{(harmonic balancing)} \end{aligned}$$

where $V_{gen}^{(n)}$ and $I_{gen}^{(n)}$ are the amplitudes of the n-th harmonic of the intrinsic drain-source voltage and current respectively, and $R_{gen}^{(n)}$ and $X_{gen}^{(n)}$ are the real and imaginary part of the load seen from the current source at the n-th harmonic frequency.

In the proposed procedure the two kinds (namely the 'harmonic balancing' and 'matching at the current source') of equations are alternately and iteratively solved until a simultaneous and consistent solution is reached: the first guess is taken from a previous analysis result, or from a linear analysis if no guess is available, as usual in non-linear algorithms. Convergence is faster than in the case of a standard harmonic balance problem: the number of equations is nearly the same (only the input matching equations are added), but the 'balancing' equations have twice-as-fast Fourier transform, and the 'matching' equations, who constitute half of the system, have an extremely fast solution. Only a few seconds are required on a 80486 PC to reach convergence in a typical case.

As an example, in fig.4 output power, large-signal gain and power-added efficiency, computed with the proposed method (solid line) are plotted *vs.* input power and compared with the results of a commercial CAD program ([11], dotted): the minor discrepancies between the two methods arise from the use of linear reactances in the proposed method. The device used is a GEC Marconi B2 medium power MESFET, whose equivalent circuit model has been extracted using both pulsed DC measurements and RF small-signal S-parameters at

multiple bias points. It must be stressed that the results of the commercial software package are obtained performing a simulated load/source pull for each input power, and the plots in fig.4 are not therefore standard power sweeps. The time needed to obtain such plots with the proposed approach is substantially smaller (more than one order of magnitude, with a non-optimized code) than that required for the simulated load/source pull.

Design procedure

For a given bias point, drive level and load resistance the optimum large-signal complex input match and reactive output match are imposed, and the performances of this 'partially optimized' stage are computed in a substantially shorter time than a single standard non-linear analysis. The method is therefore suitable for a systematic investigation and optimization of bias point, load and drive level, which cannot be automatically found: they are rather a designer's choice, subject to design compromises.

The choice of the optimum bias point (i.e. the pair of bias voltages $V_{gs,DC}$ and $V_{ds,DC}$) depends on the design specifications: if, for instance, a low-voltage design is concerned, the drain bias voltage is fixed, leaving the gate bias alone (i.e. the operating class) as a designer's choice. Similarly, if the output power must be maximized, it is usually wise to choose the drain bias voltage in order to maximize the output waveform swings, i.e. midway among physical limits (breakdown, current saturation,...).

The reactive parts of the fundamental and harmonic load terminations come out of the analysis/optimization process, as stated above. The selection of the resistive parts on the other hand requires some care: they can be used to match, as much as possible, design specifications if a 'maximum efficiency' or 'maximum power' design is concerned [4]; however, if the complexity of the harmonic-terminating circuitry is too high or if the harmonic terminations cannot effectively be controlled at such high frequencies, a good compromise has been recognized to be to short all harmonic frequencies [6, 22]. In fact, the output capacitive reactance of the active device usually dominates at higher frequencies, acting as a short circuit. Moreover, this choice leads to a great simplification of the design procedure and allows a major reduction of the number of equations to be solved (just the first four ones in the system (4)).

The resistive part of load at the fundamental-frequency only is therefore assumed in the following as a second free parameter to be optimized. The third free parameter, i.e. the drive level, can be chosen through a suitable condition. For a given bias point and load the input level can be swept from linear to saturated behavior, and the point of interest selected depending on the design specifications; possible alternatives are the 1-dB gain compression point, the maximum efficiency point or the saturated output power point. It must be noted that this procedure (source stepping) has also the well-known advantage of easing the non-linear analysis: starting from the linear solution for the first low-level point, the solution of the previous point is used as the first guess for the algorithm for any input power. This method greatly speeds up the solution, and allows a careful selection of the drive level.

In this way, only two free design parameters are now left: operating class and resistive part of the load at fundamental frequency. The proposed design procedure is based on the preliminary calculation of tables or plots of the output quantities (P_{out}, G_{LS}, η_{add}) through repeated analyses within a certain range of values of the free design parameters. Bias point (or drain-source quiescent current) is swept from class B to class A, and for each bias the load resistance is varied within a suitable range. For each pair of values the input power is selected

corresponding, in this case, to the point of maximum efficiency, as explained above. The results for a Plessey P1135 medium power MESFET are shown in fig.5 (connected symbols) as functions of the load resistance, with the quiescent current as a parameter. In the same plots, results from a commercial CAD program ([12], disconnected symbols) are shown for comparison. The two methods are in substantial agreement, but the proposed one requires about one order of magnitude less computer time.

It is easily seen that it is not possible to simultaneously maximize output power, gain and efficiency, as expected. In fact, holding the load constant and considering the variation of the output quantities with the class of operation, DC power consumption beneficially decreases moving toward class B. Nevertheless, at the same time output power and gain decrease: a point of maximum efficiency is usually present in class AB as a result of these contrasting effects.

On the other hand, for certain choices of the operating class, increasing the load resistance beyond the value corresponding to the point of maximum output power may result in a further increase of the power-added efficiency. This particular behavior is due to the increase of the large-signal gain compensating the decrease of the drain efficiency and it is typical of low-gain operating conditions (near class-B and/or at high frequencies). A compromise must then be accepted by the designer.

The analysis/optimization process can be exploited in other ways. In some applications the choice of the operating class is forced by minimum DC power dissipation (i.e. class-B) or maximum dynamic range (i.e. class A) requirements; in this case the free parameters are only the drive level and the load resistance. A quantitative trade-off, corresponding to the operating compression or back-off level and loading, must be found between maximum power and maximum efficiency conditions. For this purpose, in fig.6 constant output power and power-added efficiency contours are plotted as functions of the resistive part of the load and input power for a given bias condition. Maximum efficiency and maximum power points are not in the same region, but allow the designer to trade-off between the two optima, selecting the desired operating condition.

Conclusions

Classical approaches for the design of microwave power amplifiers have been discussed, focusing on accuracy and flexibility. More recent design procedures show indoubtedly a greater accuracy at the expense of cost and clarity of interpretation of the results. As a convenient tradeoff, a non-linear mixed analysis/optimization algorithm has been presented for the design of microwave power amplifiers. Matching conditions for maximum power and efficiency performance are imposed together with the balancing equations of the non-linear analysis in a consistent way. The algorithm performs therefore a partial source- and load-pull, yielding a pre-optimized power stage; the total computational effort however is smaller than the one required by a standard non-linear analysis for a fixed bias point and load.

The proposed method can be used as the core of a design procedure investigating the possible trade-off among operating class, drive level and optimum loading of the power stage. Comparisons with results from a commercial CAD package show a major reduction in the computational effort without substantial loss of accuracy.

Bibliography

[1]   F.N.Sechi, "Design Procedure for High-Efficiency Linear Microwave Power Amplifiers," IEEE Trans. on MTT, Vol. MTT-28, n°11, november 1980, pp.1157-1163.

[2]   S.C.Cripps, "A Theory for the Prediction of GaAs FET Load-pull Contours," 1983 IEEE MTT Symp. Dig., New York, NY, pp.221-223.

[3]   H.Kondoh,"FET Power Performance Prediction Using a Linearized Device Model," 1989 IEEE MTT Symp. Dig., pp.569-572.

[4]   D.M.Snider, "A Theorethical Analysis and Experimental Confirmation of the Optimally Loaded and Overdriven RF Power Amplifier," IEEE Trans. on Electron Devices, Vol.ED-14, N.12, December 1967, pp.851-857.

[5]   M.G.Adlerstein, M.P.Zaitlin, "Cutoff Operation of Heterojunction Bipolar Transistors," Microwave Journal, n.9, September 1991, pp.114-125.

[6]   L.J.Kushner, "Output Performance of Idealized Microwave Power Amplifiers," Microwave Journal, n.10, October 1989, pp.103-116.

[7]   F.Giannini, G.Leuzzi, E.Limiti, L.Scucchia, "Power and Efficiency Optimisation of Class-AB Narrowband Power Amplifiers," Proc. of the MIOP'93, Sindelfingen, Germany, May, pp.101-105.

[8]   M.G.Adlerstein, M.P.Zaitlin. "Power Contours for Microwave HBT," Microwave Journal, n.3, March 1993, pp.70-80.

[9]   S.Toyoda, "High Efficiency Single and Push-Pull Power Amplifiers " 1993 IEEE MTT Symp.Dig., Atlanta, GA, June 1993, pp.277-280.

[10]   F.Giannini, G.Leuzzi, E.Limiti, L.Scucchia, "Harmonic Manipulation Cure for High- Efficiency Power Amplifiers," Proc. of GAAS'94, Turin, Italy, April 1994.

[11]   LIBRA, User's Manual, EEsof Inc., Westlake Village, CA, 1993.

[12]   Microwave Design System, Hewlett-Packard Co., Palo Alto, CA, 1993.

[13]   Microwave Harmonica, Compact Software Inc., Paterson, NJ, 1993.

[14]   Microwave Tuner System, Focus Microwaves Inc., Pointe-Claire, Quebec, Canada, 1993.

[15]   Automated Tuner System, Maury Microwave Co., Cucamonga, CA, 1993.

[16]   LP1 System, ATN Microwave, Inc., Billerica, MA, 1993.

[17]   R.J.Gilmore, M.B.Steer, "Nonlinear Circuit Analysis using the Method of Harmonic Balance - A Review of the Art," Int. J. on Microwave and Millimeter-Wave Computer-Aided Engineering, Vol.1, n.1, 1991, pp.22-37.

[18]   Takayama, "A new load-pull characterisation method for microwave power transistors," 1976 IEEE MTT-S Digest, pp.218-220, June 1976

[19]   V.Rizzoli, A.Costanzo. A.Neri, "An Advanced Empirical Model for Use in Nonlinear Simulation," Proc. of the 23rd European Microwave Conference, Madrid, Spain, September 1993, pp.1103-1108.

[20]   A.Werthof, G.Kompa, "A Unified Consistent DC to RF Large Signal FET Model Covering the Strong Dispersion Effects of HEN.T Devices," Proc. of the 23rd European Microwave Conference, Madrid, Spain, September 1993, pp.1091-1096.

[21]   F.Filicori, G.Vannini, A.Mediavilla, A.Tazon, "Modeling of Deviations between Static and Dynamic Drain Characteristics in GaAs FETs," Proc. of the 23rd European Microwave Conference, Madrid, Spain, September 1993, pp.454-457.

[22]   F.Giannini, G.Leuzzi, E.Limiti, L.Scucchia, 'Iterative Design-Oriented Analysis Method for Class-AB High-Efficiency Microwave Power Amplifiers', Int. J. on Microwave and Millimeter-Wave Computer-Aided Engineering, Vol.4, n°2, April 1994.
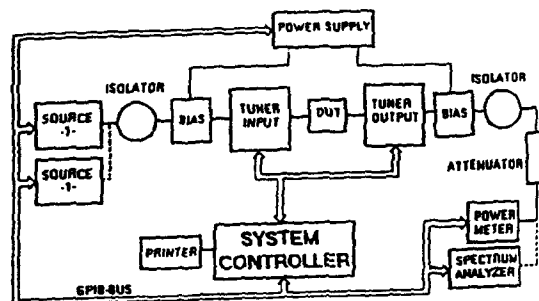
Fig.1 An example of automated
Load-Pull system

Fig.2 Equivalent circuit model of
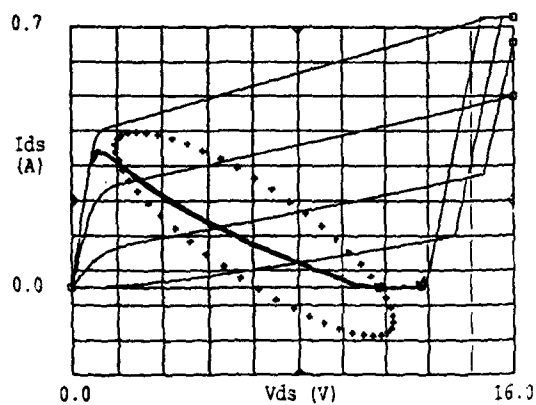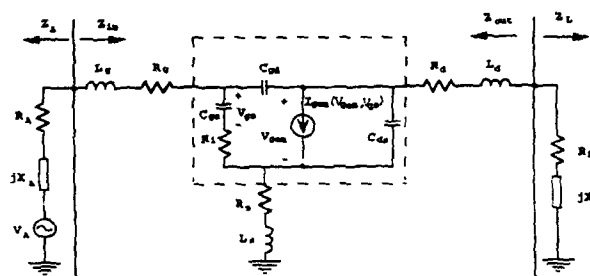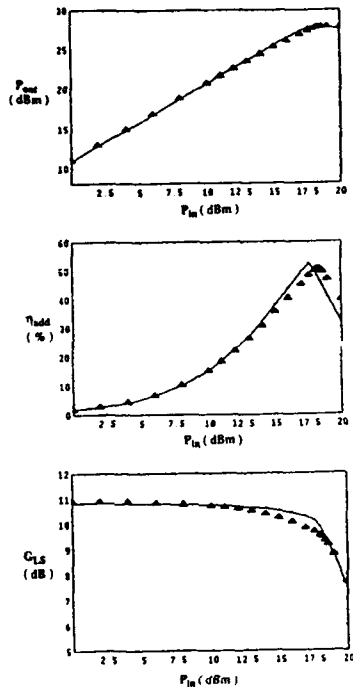an active device with the load and
source impedance indicated.

Fig.3 Typical load lines:
hysteresys-free load curve at the
intrinsic drain terminal (solid)
and at the external one (dotted).

$V_{gs,DC} = -1.75$ V

$V_{ds,DC} = 7.0$ V

$R_L = 22$ ohm

Fig.4 Output power, large-signal gain and power-added efficiency vs. input power computed with the proposed method (solid) and with a commercial CAD program ([12], dots).

Fig.5 Output power, large-signal gain and power-added efficiency *v..* **resistive part of the load, with the quiescent current as a parameter, with the proposed method (connected symbols) and with a commercial CAD program ([6], disconnected symbols).**

▲ $I_{ds,DC}/I_{dss} = 0.175$

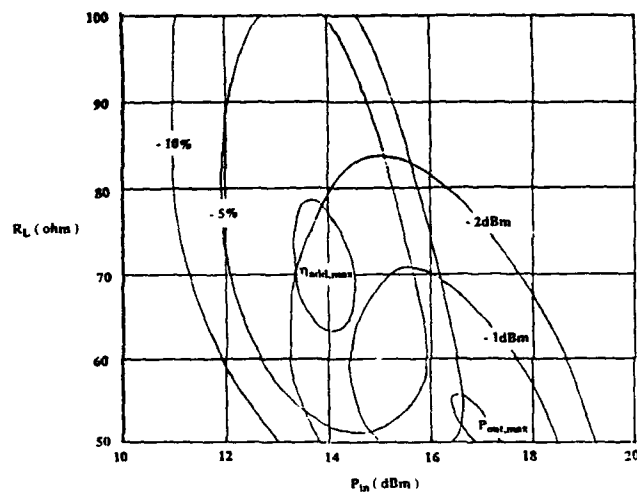● $I_{ds,DC}/I_{dss} = 0.539$
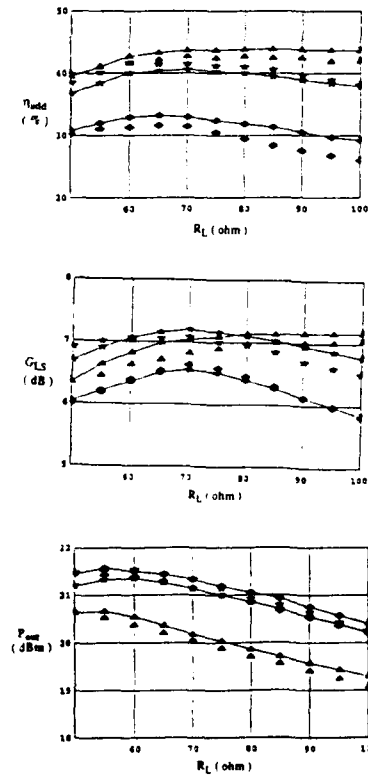
◆ $I_{ds,DC}/I_{dss} = 0.539$

Fig.6 Constant output power and power-added efficiency contours in the load/input power plane.

CHIRAL MEDIA: THEORY AND APPLICATIONS FOR MICROWAVES

Ari Sihvola[1]

Introduction

Cold War waned. We know that this fact has an enormous effect on the future of microwave technology and research. New challenges face engineers and scientists within this field, both in the East and West — if such a distinction is proper still. But the new problems we professionals are supposed to solve are not necessarily easier than those that our senior colleagues were struggling with. If military issues transformed to environmental ones, who can say that our yoke is light?

However, the future of microwaves, or any field of science for that matter, is not purely dictated by external forces. Within the field itself, the scientists can feel the pull of strong and strange attractors. Clearly, there are directions toward which the curiosity of engineers tells them to go.

One of these new, promising fields for tomorrow's microwave applications is that of new materials. Wave–material interaction has to be written anew by theoreticians, as new materials become available. Among these novel materials — or complex, or even 'exotic' materials — the most important is the class of chiral media. Chiral media, although their effects themselves are not that novel, have been studied intensively in the 1930's and 90's [1]. This interest is driven by potential applications in microwave components, like couplers and circulators, microstrip and lens antennas and radomes, and EMC and EMI applications including radar cross section reduction.

Chiral materials are handed in their geometrical structure, in other words they differ from their mirror images. This asymmetry is visible on the macroscopic level in the interaction of the medium with electromagnetic radiation, in particular in the polarization properties of the waves. As soon as the concept of polarization became treated as a scientific construct, the research on chirality also started. It is often said that this happened in the beginning of the 19th century, as Biot and Arago discovered the phenomenon of *optical activity*. However, they purely observed different classes of transparent substances, which differed in their rotatory power. The connection of this effect with the structural left–right asymmetry was not revealed until Pasteur's studies on tartaric and racemic acids in 1848.

But radiowaves and microwaves are different from optical wavelengths. So we had to wait for the discovery Hertzian waves until microwave chirality was even conceivable. In 1914 in Helsinki, the Finnish Karl F. Lindman, later professor in Turku, proved that the corresponding activity phenomenon could be observed at microwaves [2]. He synthesized a chiral medium by twisting small helices from copper wire, immersing these in cotton balls, and then positioning the balls with random orientation in a cardboard box. The length of the straight wire of the helices was 9 cm and the thickness 1.2 mm. The diameter of the spirals was 10 mm and there were 2.5 circles in one spiral. The cardboard box had length of 26 cm and the total number of spirals in the game was 700.

Lindman put his chiral box into his measurement system, shown in Figure 1. There he directs linearly polarized radio waves through the metal guide where the sample is located, and measures the linearly polarized component of the intensity of the received signal as a function of the rotation angle of the receiving antenna. His result was an

[1]Helsinki University of Technology, Electromagnetics Lab, Otakaari 5 A, FIN-02150 Espoo, Finland
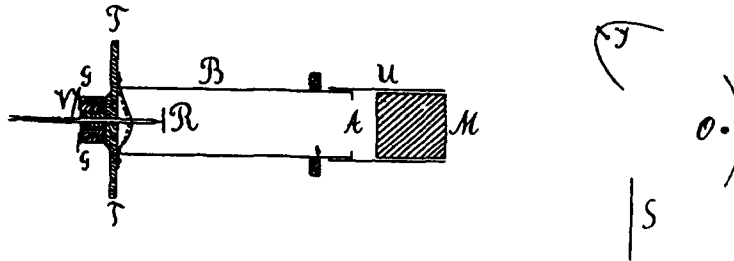
absolute confirmation of electromagnetic chirality.

Figure 1 Lindman's measuring equipment from his original article in 1914. The transmitting oscillator $\mathcal{O}$ stands in front of a reflector, which is bounded by the "standard indicator" $\mathcal{I}$ on one side, and a metal screen $\mathcal{S}$ on the other. $\mathcal{B}$ and $\mathcal{U}$ are hollow metal tubes with a circular opening $\mathcal{A}$. The sensor dipole $\mathcal{R}$ can be turned with the stick $\mathcal{T}$, and the rotation angle is read in the display formed by a pointer $\mathcal{V}$ and a protractor $\mathcal{G}$. $\mathcal{M}$ is the box where the chiral sample is put in.

Today also Lindman's way is the most common way of making artificial microwave chiral materials. How, then, to characterise electromagnetically these materials?

Electromagnetic description of chiral media

The material relations give the connection between the is the electric ($\bar{E}$) and the magnetic ($\bar{H}$) field strength, and the electric ($\bar{D}$) and the magnetic ($\bar{B}$) flux densities. The material parameters for isotropic chiral media are characterised with three scal.. parameters [3]:

$$\bar{D} = \epsilon\bar{E} - j\kappa\sqrt{\mu_0\epsilon_0}\bar{H}, \qquad \bar{B} = \mu\bar{H} + j\kappa\sqrt{\mu_0\epsilon_0}\bar{E} \tag{1}$$

The permittivity and ϵ and permeability μ measure the electric and magnetic co-polarizability of the material, and the third parameter κ contains the magnetoelectric coupling. κ is the degree of chirality. It is a dimensionless parameter. ϵ_0 and μ_0 are the permittivity and permeability of the vacuum.[2] Note that sometimes the isotropic chiral medium is called "Pasteur medium."

There are also other sets of constitutive relations in use for chiral media electromagnetics. Sometimes the chirality parameter appears as the chiral length β, sometimes as the chirality admittance ξ_c. It may be bothering to learn that in the different notations, the terms "permittivity" and "permeability" are not the same, but they denote different things! However, translation formulas are available [4] between the relations. The relations (1) are recommendable because using these, clear expressions arise for several quantities in the wave propagation analysis. Also, there is a physical limit for a the chirality parameter κ in these relations; it has to be smaller than the refractive index:

$$\kappa \leq n = \sqrt{\frac{\mu\epsilon}{\mu_0\epsilon_0}} \tag{2}$$

Electromagnetic analysis of chiral media

An important conceptual distinction in the analysis of electromagnetic fields in chiral materials is the *wavefield decomposition*. Since the medium couples electric and magnetic

---

[2]The imaginary unit j expresses the time-harmonic convention of $\exp(-j\omega t)$.

fields, it is better to look for characteristic 'eigenfields' instead of straightforwardly calculating how the 'normal' fields $\bar{E}, \bar{H}$ behave. It turns out that the Maxwell equations split into two decoupled sets if we define wavefields $\bar{E}_+$ and $\bar{E}_-$ as

$$\bar{E}_{\pm} = \frac{1}{2}(\bar{E} \mp j\eta\bar{H}), \qquad \eta = \sqrt{\mu/\epsilon} \tag{3}$$

A id furthermore, if we restrict the treatment into plane waves and homogeneous medium with no sources, the two eigenwaves propagate according to

$$\bar{E}_+(\bar{r}) = \bar{E}_+ e^{-j\bar{k}_+ \cdot \bar{r}}, \qquad \bar{E}_-(\bar{r}) = \bar{E}_- e^{-j\bar{k}_- \cdot \bar{r}} \tag{4}$$

If the direction of the propagation is defined by the unit vector $\bar{u}$, the wave vectors are

$$\bar{k}_{\pm} = \bar{u}k_{\pm} \quad \text{with} \quad k_{\pm} = k_0(n \pm \kappa) = \omega(\sqrt{\mu\epsilon} \pm \kappa\sqrt{\mu_0\epsilon_0}) \tag{5}$$

The eigenpolarizations are LCP and RCP, circularly polarized waves. One travels with a greater phase velocity than the other. But they do not couple. Microwave engineers see immediately the resemblance of this phenomenon to ferrites (or magnetoplasma) and Faraday rotation. For both media, the plane of polarization of a linearly polarized wave rotates along the propagation path. But the difference is that where ferrites are nonreciprocal and anisotropic, chiral media are reciprocal and (bi)isotropic.[3] Despite these differences, both media could be called *gyrotropic*.

The basic concept in electromagnetic analysis for fields of given electric and magnetic sources $\bar{J}, \bar{M}$ is the Green dyadic. With it the fields can be calculated in homogeneous regions:[4]

$$\begin{pmatrix} \bar{E}(\bar{r}) \\ \bar{H}(\bar{r}) \end{pmatrix} = \int \begin{pmatrix} \overline{\overline{G}}_{ee}(\bar{r} - \bar{r}') & \overline{\overline{G}}_{em}(\bar{r} - \bar{r}') \\ \overline{\overline{G}}_{me}(\bar{r} - \bar{r}') & \overline{\overline{G}}_{mm}(\bar{r} - \bar{r}') \end{pmatrix} \cdot \begin{pmatrix} \bar{J}(\bar{r}') \\ \bar{M}(\bar{r}') \end{pmatrix} dV' \tag{6}$$

The chiral-medium Green dyadic has been derived in the literature in several ways. Using the decomposition principle [3], the result looks logical:

$$\overline{\overline{G}}_{ee} = -\frac{j\eta}{2}\left[(k_+ G_+ + k_- G_-)\overline{\overline{I}} + \nabla(G_+ - G_-) \times \overline{\overline{I}} + \nabla\nabla\left(\frac{G_+}{k_+} + \frac{G_-}{k_-}\right)\right] \tag{7}$$

$$\overline{\overline{G}}_{mm} = -\frac{j}{2\eta}\left[(k_+ G_+ + k_- G_-)\overline{\overline{I}} + \nabla(G_+ - G_-) \times \overline{\overline{I}} + \nabla\nabla\left(\frac{G_+}{k_+} + \frac{G_-}{k_-}\right)\right] \tag{8}$$

$$\overline{\overline{G}}_{em} = -\overline{\overline{G}}_{me} = -\frac{1}{2}\left[(k_+ G_+ - k_- G_-)\overline{\overline{I}} + \nabla(G_+ + G_-) \times \overline{\overline{I}} + \nabla\nabla\left(\frac{G_+}{k_+} - \frac{G_-}{k_-}\right)\right] \tag{9}$$

with $G_{\pm}(\bar{r}) = \exp(-jk_{\pm}r)/4\pi r$.

In addition to these basic concepts of field analysis, several of the results and quantities needed in microwave analysis of materials have been generalized to chiral media.

---

[3] Bi-isotropic media are also called BI media. Pasteur media are one class of BI media; another subset of BI are nonreciprocal isotropic materials.

[4] Note that for the present case where there is magnetoelectric coupling, there are four dyadic components, arranged in a Green matrix.

These include reflection problems, Brewster angles, layered structures, waveguides, inhomogeneous media, etc.[5] As an example how beautifully the magnetoelectric coupling becomes visible in parametric expressions, let us take a look at the polarizability matrix of chiral sphere (Figure 2).
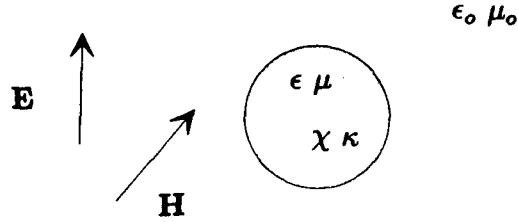

Figure 2   Chiral sphere in electromagnetic field.

The polarizability is the relation between the incident (quasistatic) field and the dipole moment induced in the sphere:

$$\begin{pmatrix} \bar{p}_e \\ \bar{p}_m \end{pmatrix} = \begin{pmatrix} \alpha_{ee} & \alpha_{em} \\ \alpha_{me} & \alpha_{mm} \end{pmatrix} \begin{pmatrix} \bar{E} \\ \bar{H} \end{pmatrix} \tag{10}$$

The polarizability relation is now in the form of a matrix, containing the co- and crosspolarizabilities

$$\alpha_{ee} = 3\epsilon_0 V \frac{(\epsilon - \epsilon_0)(\mu + 2\mu_0) - \kappa^2 \mu_0 \epsilon_0}{(\mu + 2\mu_0)(\epsilon + 2\epsilon_0) - \kappa^2 \mu_0 \epsilon_0} \tag{11}$$

$$\alpha_{mm} = 3\mu_0 V \frac{(\mu - \mu_0)(\epsilon + 2\epsilon_0) - \kappa^2 \mu_0 \epsilon_0}{(\mu + 2\mu_0)(\epsilon + 2\epsilon_0) - \kappa^2 \mu_0 \epsilon_0} \tag{12}$$

$$\alpha_{em} = -\alpha_{me} = 3\mu_0 \epsilon_0 V \frac{-3j\kappa\sqrt{\mu_0 \epsilon_0}}{(\mu + 2\mu_0)(\epsilon + 2\epsilon_0) - \kappa^2 \mu_0 \epsilon_0} \tag{13}$$

where V is the volume of the sphere. For the nonchiral, isotropic dielectric/magnetic limit $\kappa \to 0$, the polarizabilities simplify to the well-known expressions

$$\alpha_{ee} = 3\epsilon_0 V \frac{\epsilon - \epsilon_0}{\epsilon + 2\epsilon_0}, \qquad \alpha_{mm} = 3\mu_0 V \frac{\mu - \mu_0}{\mu + 2\mu_0}, \qquad \alpha_{me} = \alpha_{em} = 0 \tag{14}$$

Note here the decoupling of the electric and magnetic quantities, compared with the chiral case.

The polarizability coefficients are the basic building blocks for effective medium theories that give the connection between the structure of a mixture with its macroscopic electromagnetic parameters. In composite material design, the mixing rules are important, and with the advent of sophisticated chiral composites, it is increasingly essential to have idea how the mixing process affects the chiral properties of materials. These effects are very often nonlinear as functions of the fractional volumes of the components. Also other strange effects often appear: anomalous dispersions and percolation thresholds, for example. From the creative engineering point of view, on the other hand, all novel effects can be taken as challenges for applications that were not possible when we only had access to 'classical' materials.

---

[5] See [3] for a more detailed exposition.

Artificial chiral materials

Chirality means handedness. Or, is it more precise to say that the chiral effect arises from handed structures and objects? This distinction is by no means academic play with words, because the electromagnetic effects of a given sample of chiral medium may be totally different depending on the frequency of the wave, due to dispersion in polarization processes. Thus the rotatory effect of geometrically right-handed sample may be left-rotating, right-rotating, or may even be zero for a certain wavelength. This means that the (real part of the) chirality parameter κ may be zero for clearly chiral sample! We have to be careful to distinguish the terms 'chirality' and 'chirality parameter.'

Nature has produced chiral structures, optically active materials, those that optics scientists have studied for nearly two hundred years. Imitating Her has been the strategy of microwave engineers whose wish was to create artificial chirality. This did Lindman, and so have done his successors up to today. Figure 3 is an example. There the intention has been to have the elements randomly oriented, so that the sample would be effectively isotropic — there is no special direction. But what is the most efficient way of mixing chiral elements into a binding host matrix? Is it the explicit fabrication of vast amounts of clock springs[6] to be put together, or can we cleverly exploit polymer structures which are handed in the first place?
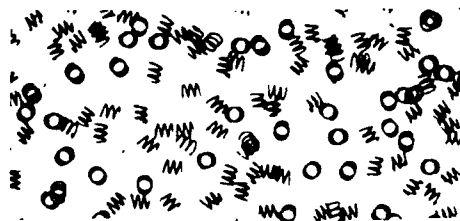

Figure 3 Artificial chiral material made by the company Finnyards Ltd. Materials Technology. The helix dimension is about 1.6 mm.

These questions are presently being studied. Nobody has yet the perfect recipe. One of the figures of merit of the chiral samples is the relative chirality parameter κ/n which has the theoretical maximum of unity. Literature does not contain too many experimental values for chirality. The following table gives, however, some idea about where we stand now.

κ	κ/n	Frequency	Reference	Note
$3.9 \cdot 10^{-5}$		549 THz	[5]	optical wave in Quartz
0.05		1.2 GHz	[2]	Lindman in 1914
0.44	0.27	10 GHz		sample of Fig. 3
1.78	0.34	15 GHz	[6]	metal helices in epoxy
0.16		15 GHz	[7]	Copper strings in dielectric
0.30	0.15	8 GHz	[8]	Ferroelectric ceramic strings

[6]Remember that although the helices appear very densely in the samples like that in Figure 3, the volume fraction of metal within the sample is fairly low. Metal phase occupies normally less or around one per cent of the total volume. But because the metal is in the form of helices, the volume of helices is considerable, and low metal volumes lead easily to situations where the helices are very close to the neighbours.

Applications

We cannot yet appreciate the wealth of applications that novel materials offer us through skillful use of microwave technology. The extra material parameter offered by isotropic chiral materials[7] can be seen as one more dimension in the space where one is looking for the optimum design, no matter what type of function one is trying to minimise or maximise.

Perhaps the most natural application of chiral materials — which are capable of rotating the polarization — is to correct depolarised fields. The degradation of polarization happens in certain lens antennas, like Luneburg and Maxwell fish-eye lenses. These inhomogeneous lenses guide the different ray paths from a dipole source into the aperture such that there the phase is even. However, due to the torsion of the geometrical optics rays, the polarization in the aperture is not uniform, and consequently the radiation efficiency is decreased.

Building the lens from chiral material, this disadvantage can be removed. The effect of chirality in geometrical optics is the rotatory power (measured with the rotation angle ϕ) integrated over the ray path s:

$$\phi = k_0 \int_{\text{ray path}} \kappa(s) ds \tag{15}$$

This angle ϕ has to compensate the polarization error. There are now several ways to fulfill the design since only the integral of the chirality function has effect. Several degrees of freedom remain.

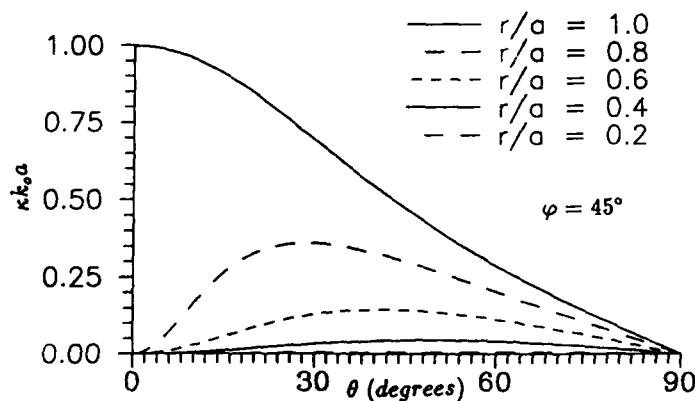As an example, Figure 4 shows one possibility of the chirality distribution along the Maxwell lens [9].

Figure 4 Normalised chirality parameter distribution of the Maxwell lens, required to correct for the polarization, as a function of the position within the lens (spherical coordinates).

Another quite interesting polarization-related application is the so called polarization transformer. This is a device which produces any desired polarization. The output is controlled by, for example the angle of the polarization of the input signal. However, this device requires slightly more general material than Pasteur medium: we need uniaxially

---

[7]And the application potential increases enormously if anisotropic chiral (bianisotropic) materials are exploited.

chiral material, which obeys.(instead of (1)) the constitutive relations

$$\bar{D} = (\epsilon_z \bar{u}_z \bar{u}_z + \epsilon_t \bar{\bar{I}}_t)\bar{E} - j\kappa \bar{u}_z \bar{u}_z \sqrt{\mu_0 \epsilon_0}\bar{H} \qquad (16)$$

$$\bar{B} = (\mu_z \bar{u}_z \bar{u}_z + \mu_t \bar{\bar{I}}_t)\bar{H} + j\kappa \bar{u}_z \bar{u}_z \sqrt{\mu_0 \epsilon_0}\bar{E} \qquad (17)$$

This type of material is conceptually as simple as isotropic chiral material, and in terms of manufacturing, it may be even simpler. There is one direction (z axis) along which the parameters are different compared to the transverse plane. Figure 5 shows the basic structure of this material.
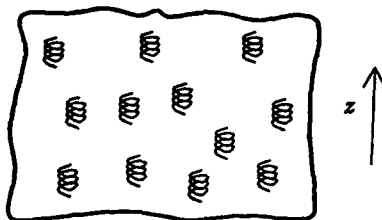
Figure 5 Uniaxial chiral medium model employing, e.g. metal helices.

It has been shown that the reflection from a planar surface of this material serves as the polarization transformer [10]. Another possibility is to have a slab which gives the same transformer effect for the transmitted wave [11].

Passive microwave components, like couplers, can be designed taking advantage of the property that the eigenwaves have different propagation constants. The coupling between open waveguides depends — in addition to the closeness of the field contact between the guides — on the phase.velocity differences of the modes. The dispersive properties of the mode velocities can be exploited in the design of narrowband filters.

Open stripline waveguides can also be used as printed circuit antennas. Chiral material may help in the antenna pattern and impedance design.

One of the main drivers for chiral microwave research has been the alleged radar cross section (RCS) reduction potential. Numerical calculations have been performed showing considerable decrease in the reflection attenuation from a metal plate using chiral coating.

Spherical *Chiroshield* is a trademark, which consists of a reciprocal chiral layer on a metal sphere [12]. The calculations of show that considerable reduction of the monostatic RCS can be achieved by including chirality on the covering layer. The backscattering can be reduced by 15 dB $\cdots$ 25 dB. In these calculations, the radius of the core is one wavelength and the Pasteur layer is one-fifth of a wavelength thick. In backscattering, lossless reciprocal chirality does not help in reducing RCS. The reason for this is that the effect of chirality is only to rotate the vector direction of the linear polarization of the propagating wave, which rotation unwinds for the backscattering case on the return path. Therefore, to produce the absorbing effect, the layer in possesses magnetic loss.

In light of the promising effects on RCS, it is understandable that chiral materials have been heavily advertised for radar cross section reduction and stealth applications. This strategy for motivating electromagnetics research projects was conspicious in the 1980's. However, objecting comments have also been raised. As was already mentioned, it is mostly the transmission characteristics, rather than reflection, that the chirality parameter of material layers affects. Therefore it may be more advantageous to concentrate on another magnetoelectric parameter for bi-isotropic media, which is the Tellegen parameter [3]. This is a measure of nonreciprocity, and it affects strongly the reflec-

tion characteristics of materials. Figure 6 shows the promising frequency behaviour of a coating made of Tellegen layer.
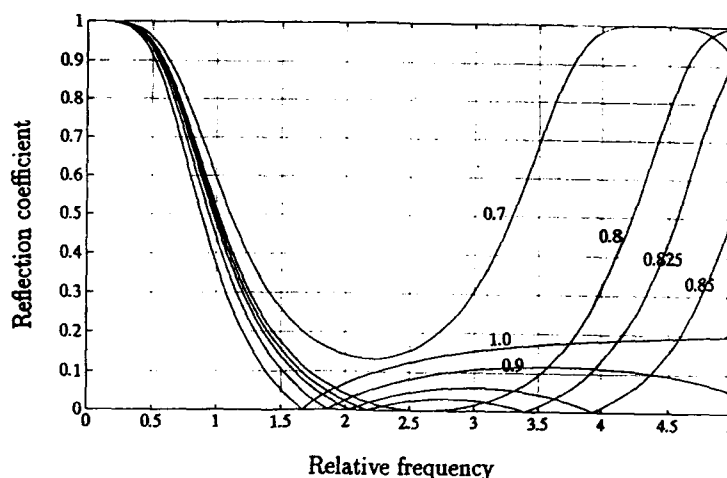

Figure 6 The copolarised reflection from a metal surface can be reduced with a nonreciprocal layer. These curves display the reflection behaviour as a function of the electrical thickness of the layer kd. The relative nonreciprocity parameter χ_r varies between 0.7 and 1. Note the extremely broad band in RCS reduction for χ_r values that are slightly larger than the normalised impedance ($\eta/\eta_0 = 0.8$) of the layer.

These ideas of how and where to apply microwave chiral and bi(an)isotropic materials have hopefully given an impression of the enormous revolution that novel media effects are creating. It is not only an additional chapter on 'strange effects' to our old electrical engineering textbooks that we need to add; this change means a total re-evaluation about the importance of all wave–material interaction aspects. It also emphasises the fact that although Maxwell equations are old, electromagnetics is by no means a sterile field. New and unexpected phenomena are there to be found.

References

[1] Proceedings of Bi-isotropics'93, Workshop on Novel Microwave Materials, *Helsinki University of Technology, Electromagnetics Laboratory Report Series,* No. 137, February 1993; Proceedings of Bianisotropics'93, International Seminar on the Electrodynamics of Chiral and Bianisotropic Media, *ibid.,* No. 159, December 1993. For a comprehensive list of references, see Unrau, U.B., "A bibliography on research in the field of bi-anisotropic, bi-isotropic, and chiral media and their microwave applications," published electronically by listserv@listserv.earn.net in the list "CHIRAL-L" on 1 February 1994. For reviews of recent chiral research, see Lakhtakia, A., "Recent contributions to classical electromagnetic theory of chiral media: what next?" *Speculations in Science and Technology,* Vol. 14, No. 1, 1991, pp. 2-17. See also, Lakhtakia, A., V.K. Varadan, and V.V. Varadan, *Time-Harmonic Electromagnetic Fields in Chiral Media,*

Lecture Notes in Physics, 335, Springer-Verlag, Berlin, 1989; Special Issue on *Wave interactions with chiral and complex media,* Vol. 6, No. 5/6 of *Journal of Electromagnetic Waves and Applications,* (Engheta, N., Guest Editor), 1992; Lakhtakia, A., "Electromagnetic theory for chiral media," in *Spiral Symmetry,* (Hargittai, I. and C.A. Pickover, Editors), pp. 281-294, Singapore, World-Scientific, 1992; Lakhtakia, A., *Beltrami Fields in Chiral Media,* Singapore, World-Scientific, 1994. Bi-isotropic media and applications, (A. Priou, Editor), in *Progress in Electromagnetics Research,* to appear in 1994, Elsevier.

[2] Lindman, K.F., "Om en genom ett isotropt system av spiralformiga resonatorer alstrad rotationspolarisation av de elektromagnetiska vågorna," *Öfversigt af Finska Vetenskaps-Societetens förhandlingar,* A. Matematik och naturvetenskaper. Vol. LVII, No. 3, 1914–1915, pp. 1-32. Lindman, K.F., "Über eine durch ein isotropes System von spiralförmigen Resonatoren erzeugte Rotationspolarisation der elektromagnetischen Wellen," *Annalen der Physik,* Vol. 63, No. 4, 1920, pp. 621-644. For a historical look on Professor Lindman's career, see Lindell, I.V., A.H. Sihvola, and J. Kurkijärvi, "Karl F. Lindman — the last Hertzian and a harbinger of electromagnetic chirality," *IEEE Antennas and Propagation Magazine,* Vol. 34, No. 3, 1992, pp. 24-30.

[3] Lindell, I.V., A.H. Sihvola, S.A. Tretyakov, and A.J. Viitanen: *Electromagnetic Waves in Chiral and Bi-Isotropic Media,* Artech House, Norwood, Mass., to appear in 1994.

[4] Sihvola, A.H. and I.V. Lindell, "Bi-isotropic constitutive relations," *Microwave and Optical Technology Letters,* Vol. 4, No. 8, 1991, pp. 295-297.

[5] Lowry, T.M., *Optical Rotatory Power,* New York, Dover, 1964, Chapter 1, Table I. Measurements made already by Biot in 1817, repeated by Lowry and Coode-Adams in 1927.

[6] Ougier, S., I. Chenerie, and S. Bolioli, "Measurement method for chiral media," *Proceedings of the 22nd European Microwave Conference,* Espoo, Finland, August 24-27, 1992, pp. 682-687.

[7] Umari, M.H., V.V. Varadan, and V.K. Varadan, "Rotation and dichroism associated with microwave propagation in chiral composite samples," *Radio Science,* Vol. 26, No. 5, 1991, pp. 1327-1334.

[8] Guérin, F., "Microwave chiral materials: a review of experimental studies and some results on composites with ferroelectric ceramic inclusions," to appear in the Special Issue of *Progress in Electromagnetics Research* on bi-isotropic media and applications, Priou, A. (editor), Elsevier, 1994.

[9] Lindell, I.V., A.H. Sihvola, A.J. Viitanen, and S.A. Tretyakov, "Geometrical optics in inhomogeneous chiral media with application to polarization correction in inhomogeneous lens antennas," *Journal of Electromagnetic Waves and Applications,* Vol. 4, No. 6, 1990, pp. 533-548.

[10] Lindell, I.V. and A.H. Sihvola, "Plane-wave reflection from uniaxial chiral interface and its applications to polarization transformer", to appear, 1994.

[11] Viitanen, A.J. and I.V. Lindell, "Uniaxial chiral quarter-wave polarization transformer," *Electronics Letters,* Vol. 29, No. 12, pp. 1074-1075.

[12] Jaggard, D.L., J.C. Liu, and X. Sun, "Spherical chiroshield," *Electronics Letters,* Vol. 27, No. 1, 1991, pp. 77-79.

STATE AND TRENDS IN TIME DOMAIN ELECTROMAGNETIC MODELLING USING THE TLM METHOD

Ulf Mueller and Adalbert Beyer*

1. Historical Background

Christiaan Huygens, a Dutch physicist, started his research on the explanation of the emission, transmission, reflection, refraction and birefringence of light in 1676 [1]. Almost at the same time Sir Isaac Newton dealt with optical research, too. In 1690 Huygens published his famous "Traité de la Lumière" [2], where he presented his theory of ether being the media for the transportation of light by mechanical interaction of its particles (see Fig. 1a) and 1b)), and therefore he declared light being a wave to be transported by elastic shocks of the ether particles.
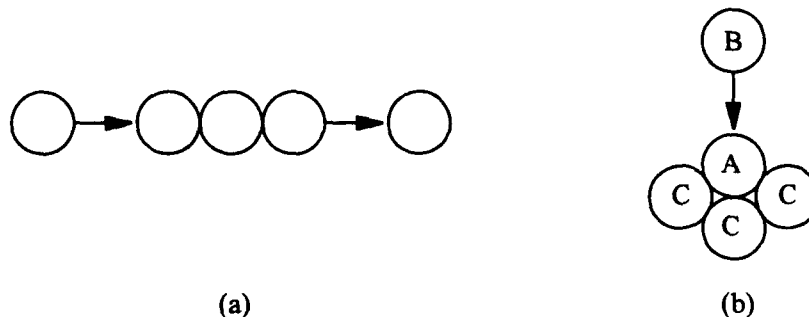

(a)   (b)

Figure 1: *(a) Huygens' explanation of the finite speed of light as evaluated by Romer as elastic shock interaction of ether particles.(b) A reproduction of the "Traité de la Lumière" to show the idea of Huygens' principle [3].*

14 years after Huygens Newton published his results on optics, explaining the light to consist of particles itself. In his opinion ether filling the whole space was not possible due to the fact that stars were moving almost without friction, as already proposed by Kepler. Henceforth, since the beginning of the 18th century there exist two totally different models of light and its propagation phenomena, which do not seem to explain the same physical background. Only with the development of the quantum theory in the middle of the 20th century these two models were found to describe the same physics. The quantum theory shows that the wave model and the corpuscular model are no concurrent descriptions but complementary aspects of the same physical phenomenon mutually supplying each others lack. Depending on the task, it is the users assignment to evaluate the arguments for choosing the very model and not the adverse.

* Duisburg University, Department of Electromagnetic Theory and Engineering, Bismarckstrasse 81, 47048 Duisburg, GERMANY

In the area of high frequency electromagnetic fields it turned out that the best description of the occurring effects is given by the wave model. This wave model together with Huygens' principle will be the main topic of the following paragraphs.

2. Huygens' principle and its implementation on a computer

In his work [2] Huygens proposed that each point in space, which is hit by a wave front, will be the source point for a wavelett of the same frequency and phase velocity as the exciting wave. The superposition of all waveletts with respect to their phases gives the resulting wave front, even after reflection, refraction or birefringence of the exciting wave. This phenomenon is represented in Fig. 2 where a wave outgoing from its source point S hit a slit (A-B) in a metal plane. The ongoing wave can be developed in a series expansion of waveletts with their envelope representing the propagating wave front.
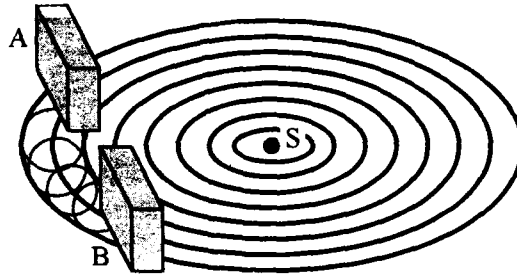
Figure 2: *Model of waves coming from sourcepoint S and hitting the metal slit A-B. The excited waveletts are depicted as well as the envelope of their superposition.*

The important development of the digital computer in the 20th century [1] led to the rediscovery of Huygens' principle as a powerful tool to describe the phenomena mentioned above. Since in the discipline of electrical engineering many problems occur associated with the distribution and propagation of electric and magnetic fields, a numerical procedure for the general solution of Maxwell's equations has to be considered incorporating Huygens' principle. Therefore, it is necessary to reformulate both in a discrete form, because digital computers can handle only discrete space and time. In order to make this principle available for the computer a four-dimensional grid of points is introduced to sample the functions $w(x,y,z,t)$, given by their Cartesian components $w(x,y,z,t) \in \{E_x, E_y, E_z, H_x, H_y, H_z\}$, representing the field components in terms of the three space and the time variables. Using different sampling rates in each dimension of the space-time domain a field component at a certain point $\vec{p}(i,j,k)$ in space at the time $_n t$ is given by

$$_n w(i,j,k) = w(i \cdot \Delta x, j \cdot \Delta y, k \cdot \Delta z, n \cdot \Delta t) \tag{1}$$

applying a notation similar to the one proposed by Yee in [4]. The numerical algorithm utilised to determine the wave propagation by decomposing them into incident, reflected and transmitted parts has to fulfill the physical law of energy conservation as well as the continuity conditions for the fields imposed by Maxwell's equations.

The first numerical approach for electromagnetic wave propagation not only dealing with Maxwell's equations (as done by Yee [4]) but also incorporating Huygens' principle was presented by Johns and Beurle [5] in 1971. They simply proposed a perpendicular mesh of interconnected transmission lines to fit the above mentioned requirements for the two-dimensional case. The differential equations for the voltages and currents on the network of linear, homogeneous transmission lines and the Maxwell's equations describing the wave propagation in a linear, homogeneous medium are built up analogously. Johns and Beurle found out that this analogy meets the discretized Huygens' principle additionally.
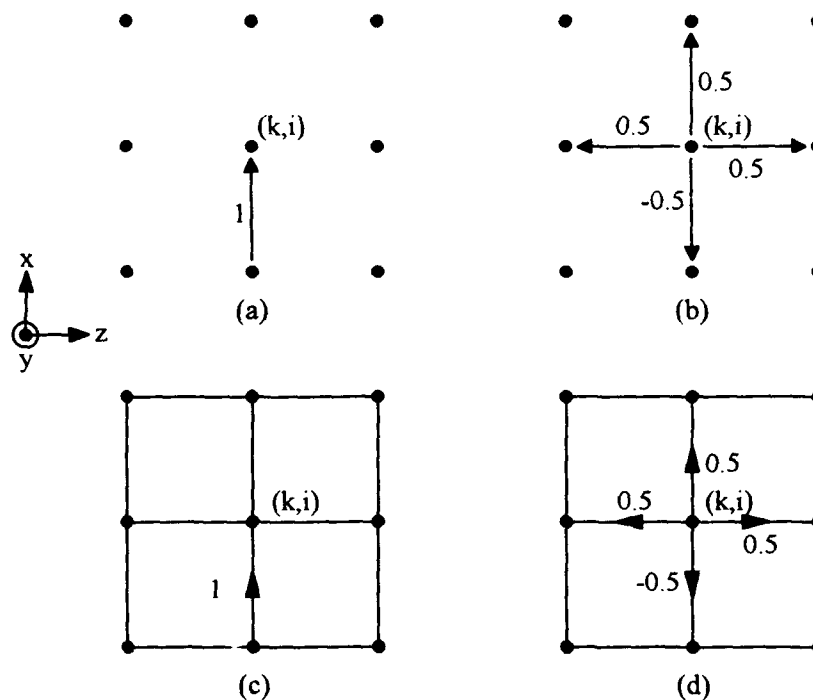

Figure 3: *Two dimensional discrete wave scattering as imposed by the Huygens' principle: (a) shows the incident wave of normalised electrical field amplitude 1 and (b) reflected and transmitted parts of the wave. (c) depicts the incident normalised voltage on the network of interconnected transmission lines and (d) the reflected and transmitted voltage waves.*

Fig. 3 shows the behaviour of the discretized Huygens' principle for incident, transmitted and reflected parts of the electrical field (Figs. 3a) and 3b)) in comparison to the reaction of the transmission line network (Figs. 3c) and 3d)) induced by an incident normalised voltage. It can clearly be seen that both descriptions have identical responds. The incident, transmitted and reflected parts in both cases have the same amplitude fulfilling the requirement of the field continuity condition by th difference in sign while the energy (proportional to the square of the amplitudes) split into four identical parts.

Now, consider a point $\vec{p}(k,i)$ in the two-dimensional space and superpose its reaction on the four possible directions of incident waves. Henceforth, the equations

$$
\begin{pmatrix} {}_nE_y^r(k+0,i) \\ {}_nE_y^r(k-0,i) \\ {}_nE_y^r(k,i+0) \\ {}_nE_y^r(k,i-0) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} {}_nE_y^i(k+0,i) \\ {}_nE_y^i(k-0,i) \\ {}_nE_y^i(k,i+0) \\ {}_nE_y^i(k,i-0) \end{pmatrix}
\tag{2}
$$

and

$$
\begin{pmatrix} {}_nU_1^r(k,i) \\ {}_nU_2^r(k,i) \\ {}_nU_3^r(k,i) \\ {}_nU_4^r(k,i) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} {}_nU_1^i(k,i) \\ {}_nU_2^i(k,i) \\ {}_nU_3^i(k,i) \\ {}_nU_4^i(k,i) \end{pmatrix}
\tag{3}
$$

can be determined to describe the scattering behaviour of the discrete space (Eq. (2)) and the transmission line network (Eq. (3)) in terms of incident and reflected electrical fields and voltages, respectively. These equations yield the generalised scattering formulation of an n-port by scattering matrix $\tilde{S}$ and vectors of incident waves $\vec{a}$ and reflected waves $\vec{b}$ as given by

$$
\vec{b} = \tilde{S} \cdot \vec{a}.
\tag{4}
$$

In Fig. 4 the equivalent network of the so-called "shunt node" is given. Each quarter of the node in Fig. 4a) is replaced by a series connection of an inductance L and a capacitance C as demonstrated in Fig. 4d). The result is depicted in Fig. 4b) leading to a total capacitance of 4C taking the shunt connection of the four node parts into account. This model is only valid as long as the highest evaluated frequency propagating on the mesh has a wavelength λ much larger than the mesh parameter Δh, hence the variation of the fields over the distance Δh is negligible. Fig. 4b) gives the dimensions of the shunt node as well as the numbering of the ports with respect to the coordinate axes.
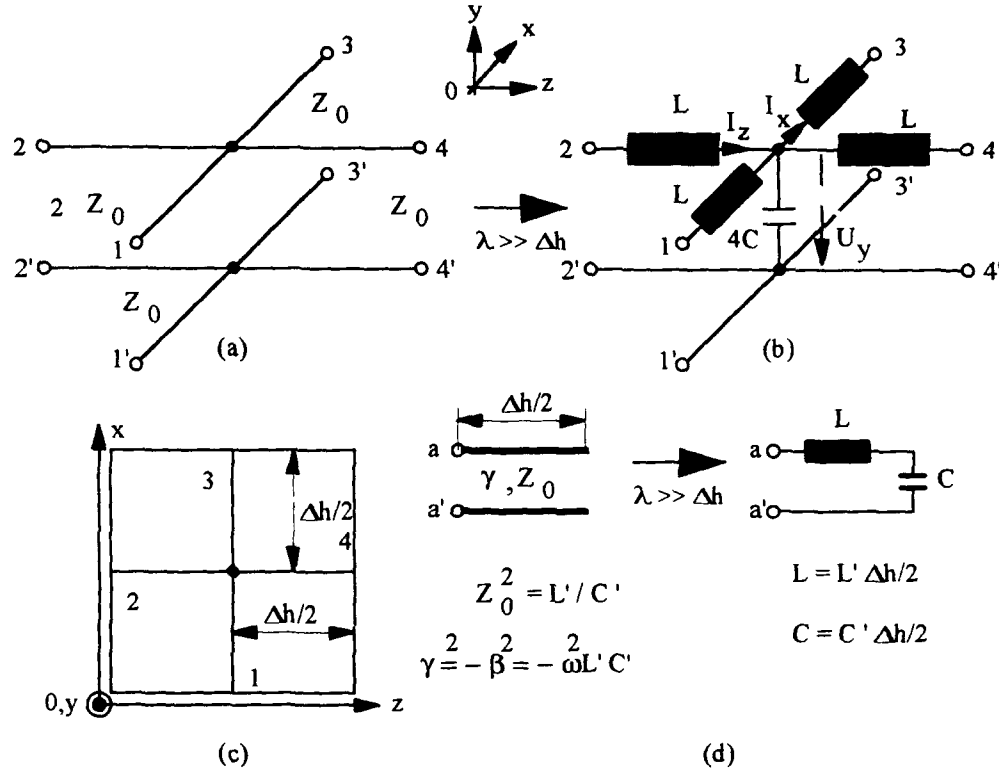
Figure 4: *The interconnected transmission lines in (a) can be replaced by a lumped element network as given in (b) as long as the wavelength of the highest excited frequency is much larger than the discretization parameter Δh, means $\lambda >> \Delta h$. The equivalence for one quarter of the node is given in (d) and (c) shows the numbering of the ports at the node with respect to the coordinate axis.*

Introducing the voltage drop U_y over the capacitance and the currents I_z and I_x in the direction of the corresponding coordinate axes, these system variables yield the following system of linear partial differential equations

$$\frac{\partial U_y(k,i)}{\partial x} = -L' \cdot \frac{\partial I_x(k,i)}{\partial t}$$

$$\frac{\partial U_y(k,i)}{\partial z} = -L' \cdot \frac{\partial I_z(k,i)}{\partial t} \tag{5}$$

$$\frac{\partial I_z(k,i)}{\partial z} + \frac{\partial I_x(k,i)}{\partial x} = -2C' \cdot \frac{\partial U_y(k,i)}{\partial t}.$$

The Maxwell's equations for $TH_{nm,y}$- or $TE_{n0,z}$-waves lead to

$$\frac{\partial E_y(k,i)}{\partial x} = -\mu_0 \cdot \frac{\partial H_z(k,i)}{\partial t}$$

$$\frac{\partial E_y(k,i)}{\partial z} = \mu_0 \cdot \frac{\partial H_x(k,i)}{\partial t} \qquad (6)$$

$$\frac{\partial H_x(k,i)}{\partial z} - \frac{\partial H_z(k,i)}{\partial x} = \varepsilon_0 \cdot \frac{\partial E_y(k,i)}{\partial t}$$

as the system of linear partial differential equations for the electrical and magnetic fields.

Therefore, the equivalencies

$$E_y \hat{=} U_y \; ; \; H_z \hat{=} I_x \; ; \; H_x \hat{=} -I_z$$

$$\mu_0 \hat{=} L' \; ; \; \varepsilon_0 \hat{=} 2C' \qquad (7)$$

can be recognized between the voltage and the electrical field, the currents and the magnetic field components as well as for the material properties μ_0, ε_0 and the capacitances C' and inductances L' per unit length of the network model [6].

Johns and Akhtarzard also found, that a series connection of small transmission line pieces [7] also fit the requirements of the discrete Huygens' principle . Applying the Babinett's principle [8], which describes the dualism of electric and magnetic field, voltage and current, and impedance and admittance, to this series connection give the equations identical to those of the shunt node. Hence, it is not necessary to introduce both types of nodes. This would only be necessary in order to switch from two-dimensional modelling to three-dimensional modelling, which requires to solve the three-dimensional vector wave equation

$$\frac{\partial^2 \vec{X}(x,y,z,t)}{\partial x^2} + \frac{\partial^2 \vec{X}(x,y,z,t)}{\partial y^2} + \frac{\partial^2 \vec{X}(x,y,z,t)}{\partial z^2} - \frac{1}{v_{ph}^2}\frac{\partial^2 \vec{X}(x,y,z,t)}{\partial t^2} = \vec{0} \qquad (8)$$

in a Cartesian coordinate system with $\vec{X} \in \left\{ \vec{E}, \vec{H} \right\}$.

This can be done in two different ways: either by a combination of series and shunt nodes [9] with distributed voltages and currents identical to the distributed field components as in Yee's finite difference time domain (FDTD) mesh [4], or by the symmetrical condensed node (SCN) derived by Johns [10] in 1987, describing all six field components at the same point in space.

The ports of the SCN numbered 1 to 12 are for the simulation of free space [10], whereas the ports numbered 13 to 18 (s. Fig. 5) are for the modelling of material properties or irregular graded meshes [11]. The relation between the incident and the reflected voltages at the ports, numbered as given in Fig. 5 and orientated with respect to the direction of the coordinate axes can once again be given by a scattering equation similar to Eq. (4). Opposing the shunt node the vectors of incident and reflected voltages for the symmetrical condensed node consist of 18

elements each, while the scattering matrix is a set of 324 elements building a sparse square matrix [10] of order 18.
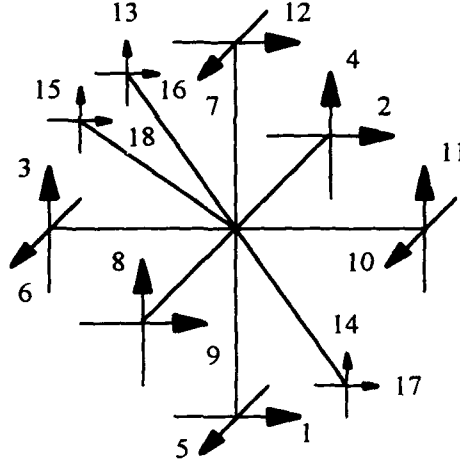

Figure 5: *The symmetrical condensed node as proposed by Johns [10] in 1987.*

Up to now we discussed only the relationships between incident and reflected voltages at the ports of the chosen node, but the TLM algorithm consists of a second part, the exchange. The exchange bases on the fact, that the voltages reflected from the ports of a node at a certain time are incident ones to the neighbouring nodes at the next time step after travelling along a connecting transmission line of mesh parameter Δh length.

3. Boundary Conditions

This exchange part of the algorithm and the fact, that the synchronism of pulses have to be guaranteed during the simulation with respect to the discrete time, impose the boundaries only to be placed halfway between the nodes. Henceforth, boundaries have to be specified in the way they interact with the electric and magnetic fields halfway between the points, where they are even known. This seemed to be a problem, but since on a transmission line the equations

$$U = U^i + U^r \qquad \text{and} \qquad I = \frac{U^i - U^r}{Z_0} \qquad (9)$$

are valid, whereby the voltage U represents the total electric field, while the current I represents the total magnetic field, the boundary conditions can be specified by simply inserting a reflection coefficient of $r = -1$ for an electric wall and $r = +1$ for a magnetic wall. These reflection coefficients lead to

$$\vec{E} \times \vec{n} = \vec{0}\Big|_{Boundary} \qquad \text{and} \qquad \vec{H} \cdot \vec{n} = 0\Big|_{Boundary} \qquad (10)$$

-154-

for an electric wall, while for a magnetic wall the electric and magnetic fields yield

$$\vec{E} \cdot \vec{n} = 0\big|_{Boundary} \qquad \text{and} \qquad \vec{H} \times \vec{n} = \vec{0}\big|_{Boundary} \qquad (11)$$
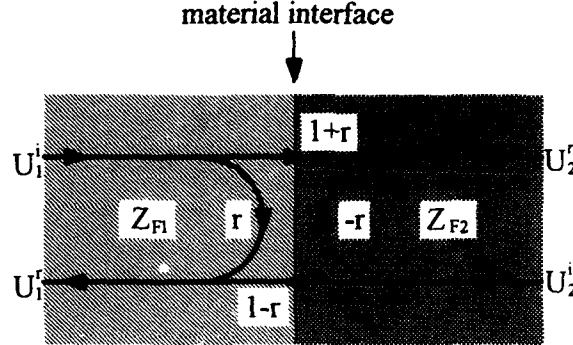
material interface

Figure 6: *Voltage wave exchange at a material interface [12] for the two-dimensional TLM model*.

At a material interface as given in Fig. 6 for the two-dimensional case the simple reflection coefficient r extends to the full scattering equation

$$\begin{pmatrix} U_1^r \\ U_2^r \end{pmatrix} = \begin{pmatrix} r & 1-r \\ 1+r & -r \end{pmatrix} \cdot \begin{pmatrix} U_1^i \\ U_2^i \end{pmatrix} \qquad \text{with} \qquad r = \frac{Z_{F2} - Z_{F1}}{Z_{F2} + Z_{F1}} \qquad (12)$$

between the incident voltages and the reflected voltages with respect to the wall, because the incident and reflected voltages act like the electric field components tangential to the boundary fulfilling the continuity conditions of the electric and magnetic fields. Written in terms of a matrix equation Eq. (12) gives

$$\vec{U}^r = \vec{r} \cdot \vec{U}^i \qquad \text{with} \qquad \vec{r} = \begin{pmatrix} r & 1-r \\ 1+r & r \end{pmatrix} \qquad (13)$$

for the vectors of the incident and reflected voltages $\vec{U}^i = \begin{pmatrix} U_1^i & U_2^i \end{pmatrix}^T$ and $\vec{U}^r = \begin{pmatrix} U_1^r & U_2^r \end{pmatrix}^T$, respectively. In the three-dimensional case there exist two tangential field components represented by the two vectors $\vec{U}_a$ and $\vec{U}_b$ for the perpendicular polarisations, and for each of them an equation similar to Eq. (13) can be determined. The superposition for both perpendicular polarisations then gives:

$$\begin{pmatrix} \vec{U}_a^r \\ \vec{U}_b^r \end{pmatrix} = \begin{pmatrix} \vec{r} & \vec{0} \\ \vec{0} & \vec{r} \end{pmatrix} \cdot \begin{pmatrix} \vec{U}_a^i \\ \vec{U}_b^i \end{pmatrix} \qquad (14)$$

This boundaries enable us to introduce a first type of problems solvable with the TLM method, the eigenvalue problem.

4. State of the art

4.1 Eigenvalue Problems (Boundary Value Problem)

The eigenvalue solution using TLM is based on a formulation of the structure under study shielded by a combination of electric and magnetic walls. Initially losses are not taken into account in order to conserve the energy. The walls have to be placed apart from the structure, far enough to declare their influence to be negligible, but near enough to guarantee the waveguide not to be overmoded.

After raising a wave, it propagates through the structure. The reflected parts interfere with the propagating wave in a way that the field distributions corresponding to the eigenvalues are superposed, while all other parts decrease by elimination. The total energy inserted into the system remains constant, as long as no losses are considered.

The eigenvalue problem can be solved by Fourier transformation of the impulse response obtained at an observation point. Since the time signal is a limited series [12] of N discrete pulses

$$X(t) = \sum_{n=1}^{N} {}_{n}X \cdot \delta(t - n\Delta t) \tag{15}$$

the Fourier integral

$$\mathfrak{I}\{X(t)\}(f) = \int_{-\infty}^{+\infty} X(t) \cdot e^{-j2\pi f t} \mathrm{dt} \tag{16}$$

is reduced to the two series

$$\mathrm{Re}\{\mathfrak{I}\{X(t)\}(f)\} = \sum_{n=0}^{N} {}_{n}X \cdot \cos(2\pi \, nf\Delta t)$$

$$\mathrm{Im}\{\mathfrak{I}\{X(t)\}(f)\} = -\sum_{n=0}^{N} {}_{n}X \cdot \sin(2\pi \, nf\Delta t). \tag{17}$$

for the real and the imaginary part of the Fourier transform. Defining the modulus of $\mathfrak{I}\{X(t)\}(f)$ to be

$$\|\mathfrak{I}\{X(t)\}(f)\| = \sqrt{\mathrm{Re}^2\{\mathfrak{I}\{X(t)\}(f)\} + \mathrm{Im}^2\{\mathfrak{I}\{X(t)\}(f)\}} \tag{18}$$

the eigenfrequencies f_{eigen} corresponding to the eigenvalues can be evaluated from the spectral function by

$$\lim_{f \to f_{eigen}} \|\mathfrak{I}\{X(t)\}(f)\|^2 \longrightarrow \|\mathfrak{I}\{X(t)\}(f)\|^2_{max}. \tag{19}$$

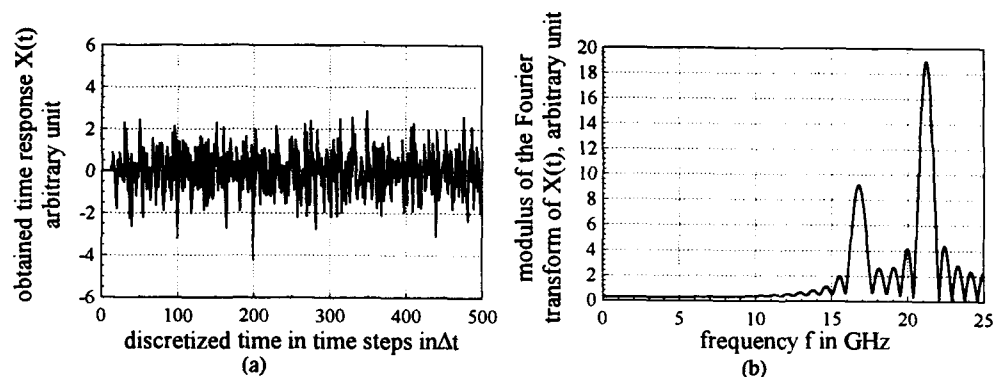whereby X(t) represents any selected field component at the point of observation.

Figure 7: *Time series X(t) of an eigenvalue analysis (a) and the corresponding spectral function after Fourier transformation (b).*

Fig. 7 gives an example of a time series (Fig. 7a)) and the corresponding Fourier transform (Fig. 7b)). Note that the frequency axis in Fig. 7b) depicts the frequency in the real structure found by multiplying the TLM network frequency by $\sqrt{2}$, which must be done due to the fact the discrete system exhibits numerical dispersion. The Fourier transform shows clearly a response different to a line spectrum but superposed by a $\left(x^{-1} \cdot \sin(x)\right)$-function due to the inevitable truncation of the impulse response. If the number of time steps is too small, the lobes get wider and wider until they interfere with each other. It has to be mentioned, that like in real field measurement, the position of input and output points as well as the nature of the selected field components will affect the magnitudes of the spectral lines. For instance, if the output nodes are situated close to a minimum of a particular mode, the eigenfrequency corresponding to this mode will not appear in the frequency response, which can be used to either suppress or enhance certain modes [11].

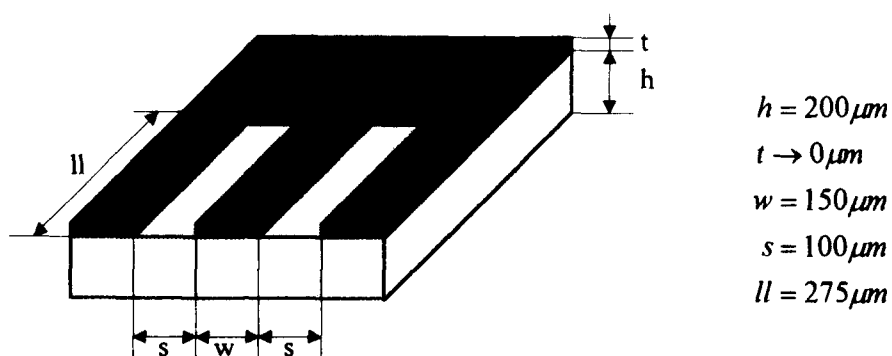As an example for the eigenvalue formulation let us consider the coplanar short in Fig. 8.

$h = 200 \mu m$

$t \rightarrow 0 \mu m$

$w = 150 \mu m$

$s = 100 \mu m$

$ll = 275 \mu m$

Figure 8: *Dimensions of the coplanar short under study.*

The knowledge of the frequency behaviour of a coplanar short is of interest especially for calibrating measurement systems, because there is a deviation between the real and the perfect coplanar discontinuity, due to the field distortion at the end of the coplanar line. The properties of the coplanar short in Fig. 8 are calculated from the eigenvalues of the resonant structure with electric walls around it, except the front wall, which is either an electric or a magnetic wall assuming a half- or quarter-wave resonator, respectively. Therefore, at first the apparent dielectric constant of the coplanar waveguide itself is calculated from a resonant arrangement for several frequencies. Thereafter, the eigenvalues of the resonator containing the short are obtained.
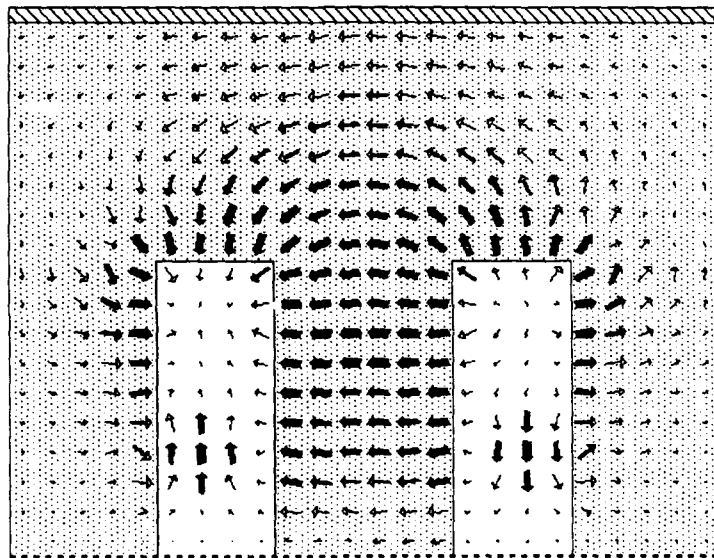

Figure 9: *Field distribution of the magnetic lines of force tangential to the upper dielectric surface of the short [13].*

The deviation between the real and an ideal short can be seen in a distortion of the field distribution of the electromagnetic field. This is depicted in Fig. 9 where the tangential magnetic field on the upper dielectric surface of the dielectric substrate is plotted, leading to a current density vector perpendicular to the lines of magnetic force. In despite to the often shown time domain field patterns a discrete Fourier transformation (DFT) was applied to each node in order to evaluate the field distribution for a certain eigenfrequency (here about $f_{eigen} \approx 50$ GHz). The distortion of the field results in an extension of the flow pattern into the short metallization. Due to this extension the behaviour is usually described by a virtual elongation Δl of the coplanar line, terminated with an perfect short. But due to the massive concentration of magnetic field energy in the area of the short metallization, also a description by an inductance added to the coplanar line seems to be sensefull.
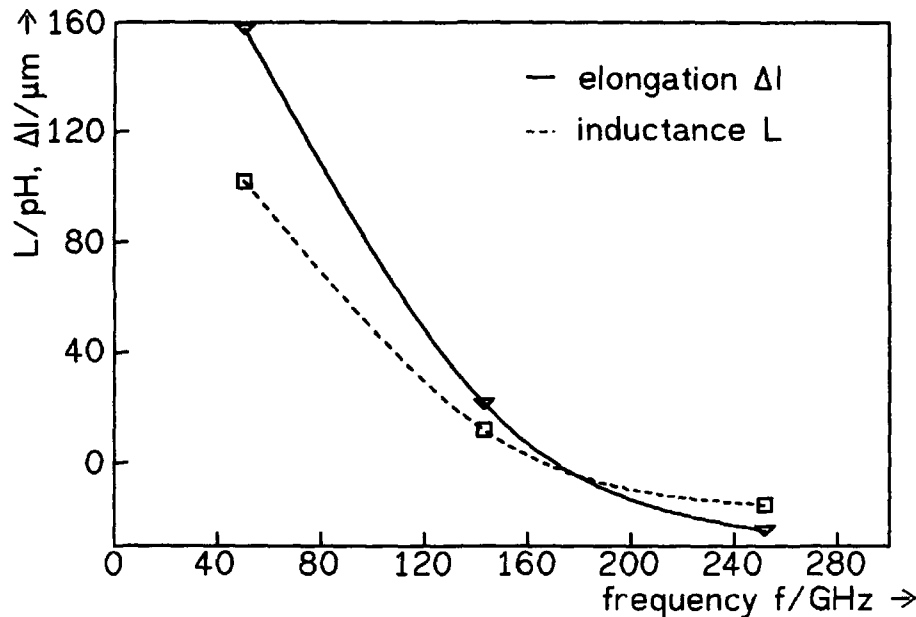
Figure 10: *Virtual transmission line elongation Δl and equivalent inductance L versus frequency f [13].*

Fig. 10 depicts the virtual elongation Δl as well as the equivalent inductance L versus the frequency f. The change in sign of the inductance L shows, that the distortion of the field at the coplanar short can not only be described as an inductance for higher frequency ranges but requires a refined model.

This example has shown the applicability of the eigenvalue formulation to microwave structures, but the evaluation requires an enormous numerical expense. Several attempts have been made to overcome this immense numerical effort, without increasing the truncation error. Saguet and Pic proposed the possibility of weighting the limited time series by different types of special tools, like Hamming or Hanning windows [14] leading to a convolution with the Fourier transform of the corresponding window function in the frequency domain, and therefore, sharper peaks were obtained. Further, Wills [15] used Prony's method of special series expansions in the frequency domain to obtain the spectral function. Nevertheless, the lack of knowledge of the transfer function for the whole frequency range under study remains. Therefore, a second treatment has been developed, which can be used to determine the frequency characteristics of a structure for a whole frequency range by a single computation run. This second area of TLM application, the so-called transient type analysis or initial boundary value problem, is focused on in the following.

4.2 Transient Analysis (Initial Boundary Value Problem)

Opposing the method mentioned above, where steady states were assumed, we now consider a time limited signal for efficient analysis of the scattering behaviour of microwave and millimeter wave structures. Therefore, we first focus on Fig 11, where a some results of a transient type analysis at several time steps for a Gaussian pulse travelling upon two step-connected microstrip lines are shown. The electrical field distribution of the field component perpendicular to the upper dielectric surface is plotted.
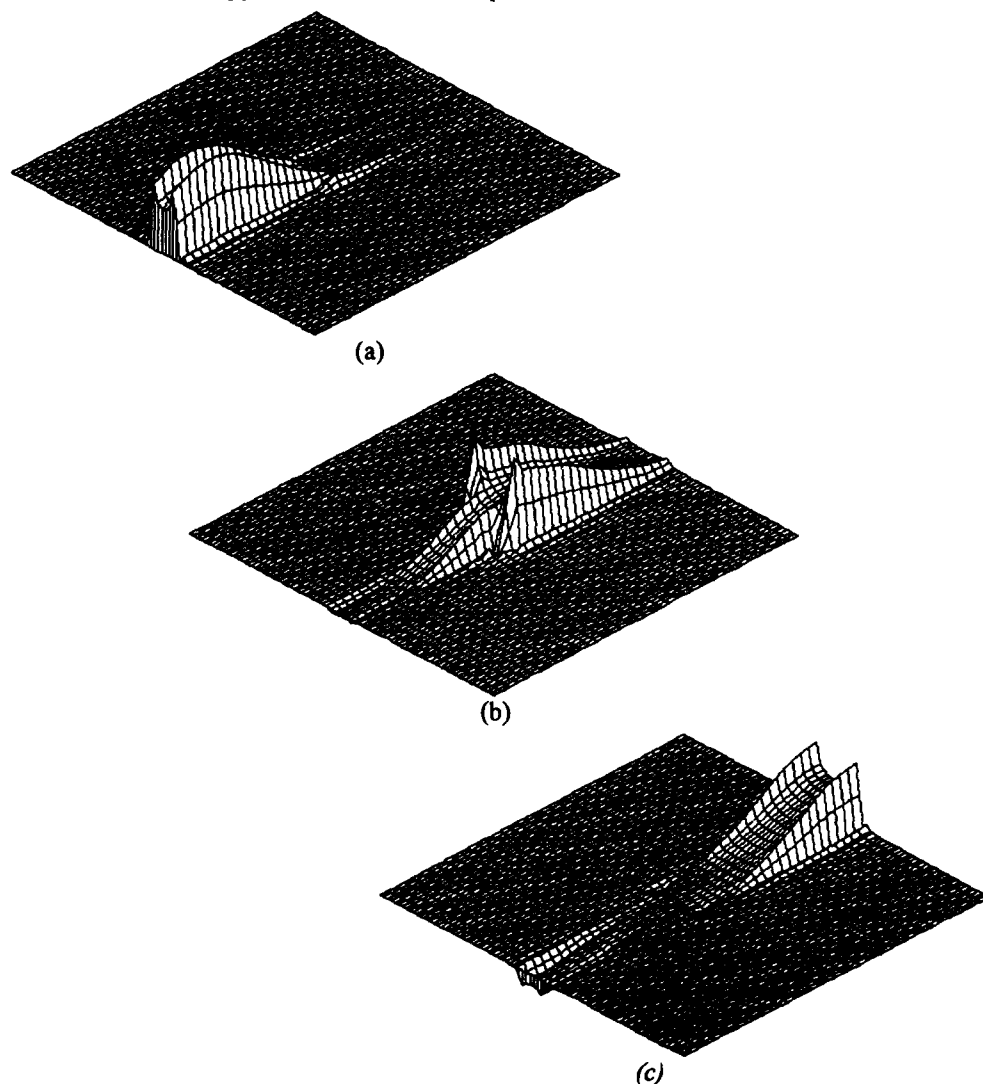
(a)

(b)

(c)

Figure 11: *A transient type analysis at different times [16]; the distribution of the electric field component perpendicular to the upper dielectric surface of a microstrip step is shown: (a) the exciting pulse on the smaller line in front of the microstrip step, (b) the maximum of the pulse at the step and (c) the reflected and transmitted pulses after scattering at the step.*

First the propagating Gaussian pulse on the smaller line is depicted (Fig. 11a)). In the second picture, the maximum of the Gaussian pulse has reached the step (Fig. 11b)), while in the last picture (Fig. 11c)) the transmitted and reflected parts of the propagating wave can clearly be detected. Trying to characterise the fundamental mode of the microstrip waveguide by a voltage integral from the strip to the ground plane, the voltages versus time at the ports can be obtained in terms of incident, reflected and transmitted voltage waves as given in Fig. 12.
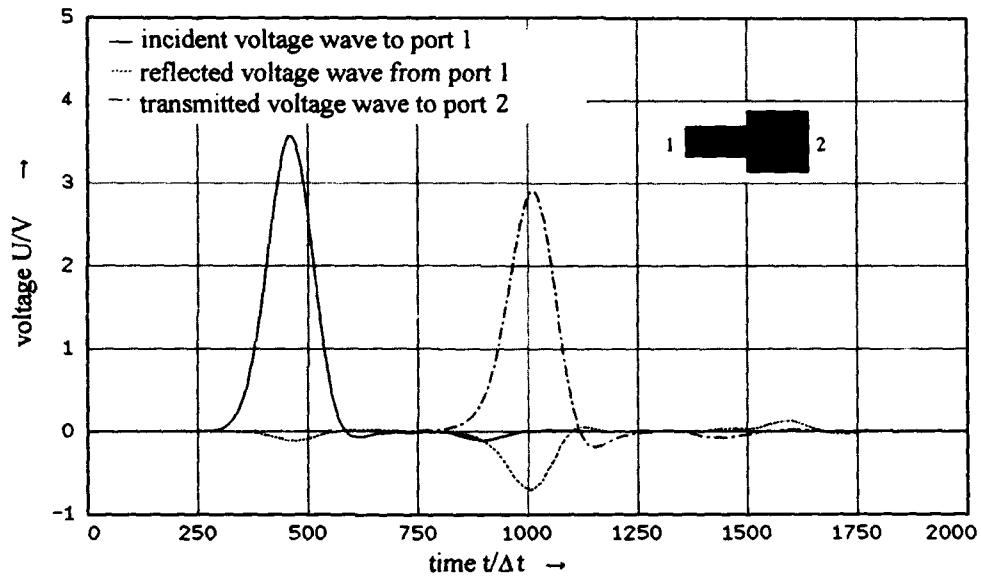

Figure 12: *Typical voltage response of a transient analysis: the voltage response of a microstrip step, incident: reflected and transmitted parts are shown [16].*

After transforming these time series into the Fourier domain the scattering parameters are calculated by simply dividing the Fourier transforms, e.g. the scattering parameter determining the energy transmitted from port k to port i yield

$$s_{ik}(f) = \sqrt{\frac{Z_{0,k}(f)}{Z_{0,i}(f)}} \cdot \frac{\Im\{U_i(t)\}(f)}{\Im\{U_k(t)\}(f)}, \tag{20}$$

whereby $Z_{0,x}(f)$, (x=i,k) is denoted to be the frequency dependent wave impedance at a port x, (x=i,k) and $U_x(t)$, (x=i,k) is the time series for the voltage integral at this port x, (x=i,k). The results are given in Fig. 13a) for the absolute value and in Fig. 13b) for the phases of s_{11} and s_{21} for the microstrip step. The agreement between the TLM analysis and results obtained from a magnetic wall model is good. The small deviations in the phases of s_{11} and s_{21} occur due to wrong positioning of the so-called absorbing boundaries, which additionally have to be introduced in order to extend the computational domain virtually to infinity. State of the art is to use a simple reflection coefficient [12] for dominant mode termination.
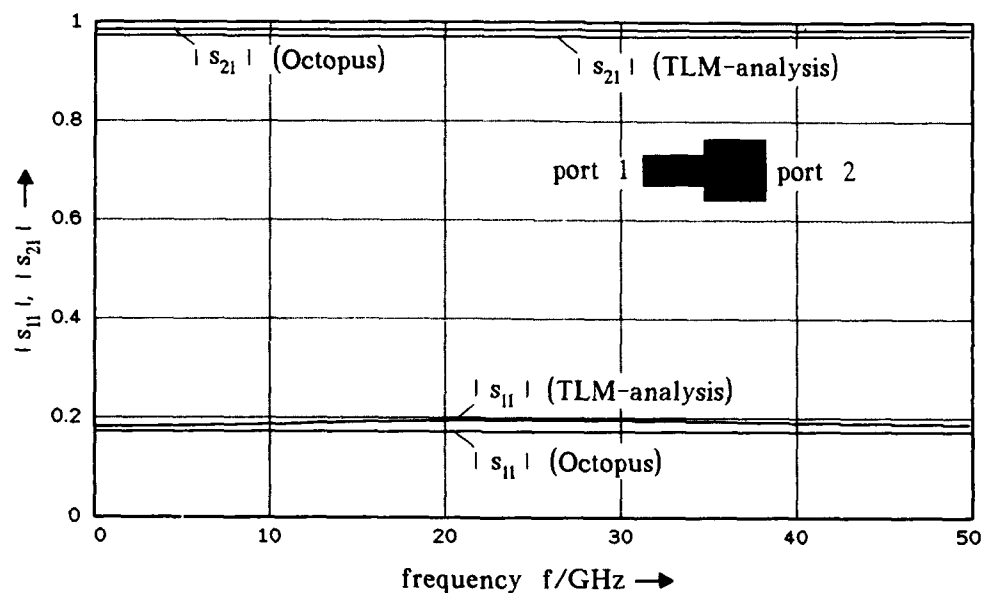
(a)
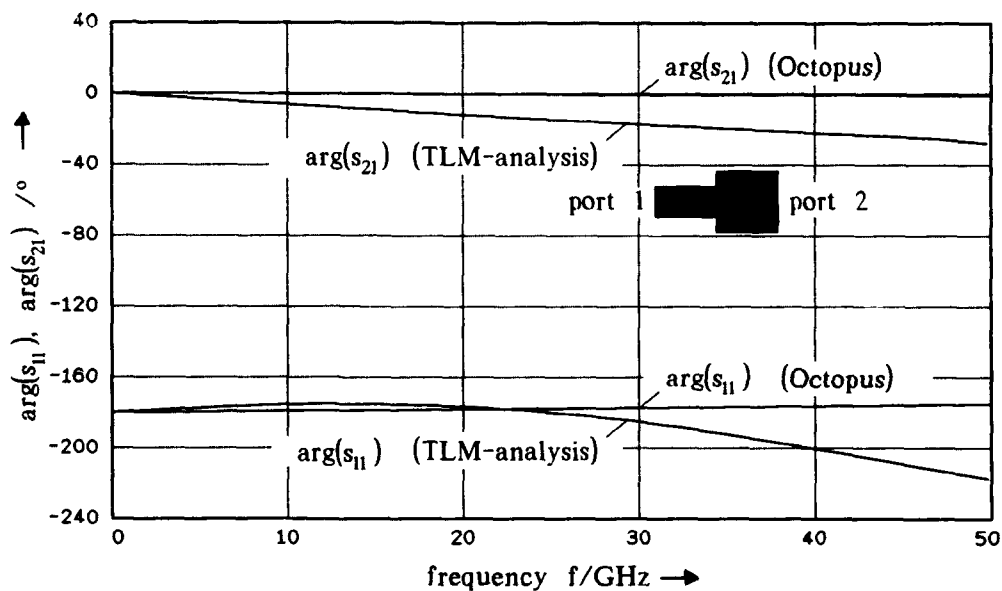
(b)

Figure 13: *Scattering parameters of a microstrip step evaluated by the TLM method [16] in comparison to a magnetic wall model [17]: (a) absolute value of s_{11} and s_{12} versus frequency and (b) angles of s_{11} and s_{12} versus frequency.*

5. Trends in research activities

5.1 Absorbing boundaries

Many papers on the TLM method are presenting results evaluated by the eigenvalue formulation using 2D or 3D meshes to determine the eigenvalues of the structures under consideration [5,9]. Since Johns proposed the use of a SCN [10] for the simulation of electromagnetic fields, the efficiency of the TLM method has steadily approached that of the FDTD method because of its favourable dispersion characteristics and the yield of six field components at one point in space. When employed in a transient time regime this type of node is very efficient in analysing complex structures [11]. In contrast to the resonant cavity formulation, the transient type analysis requires a larger computational domain in order to simulate the infinite space. Therefore, several types [11,18,19,20,21] of the so-called absorbing boundary conditions (ABCs) are used to simulate a virtual extension of the limited computational domain and reduce the required memory and simulation time. Each of them strives for a solution of Sommerfeld's radiation condition

$$\lim_{|\vec{r}| \to \infty} |\vec{r}| \left(\frac{\vec{r}}{|\vec{r}|} \times \vec{H}(\vec{r}) + \frac{\vec{E}(\vec{r})}{Z_F} \right) \to \vec{0} \tag{21}$$

with the characteristic field impedance Z_F and the vector $\vec{r}$ denoting the distance between the source point and the observation point, by applying special algorithms to either the electrical or the magnetic field or to both at the same time. They usually predict the behaviour of the electromagnetic field at a certain boundary point from values known in front of it for a first order approach, and beside of it and several steps in front of it for higher order approximations. One possibility is to utilize the properties of the one-dimensional wave propagation. The wave equation (Eq. (8)) in terms of Cartesian components can be solved under the assumption of plane wave propagation in positive x-direction by

$$E_{\tan}(x,t) = E_{\tan}(t - \frac{x}{v_{ph}}) \tag{22}$$

where $E_{\tan} \in \{E_y, E_z\}$ is the electrical field strength transversal to the propagation direction (tangential to the boundary) and v_{ph} is the phase velocity. The absorbing boundary algorithm used for the fields at the boundary $x_b = I \cdot \Delta h$ at the time $t_b = n \cdot \Delta t$ assumes that it is the same as just one discretization step Δh in front of the boundary at an earlier time $t_b - t_0$ which demands

$$E_{\tan}(t_b, x_b) = E_{\tan}(t_b - t_0, x_b - \Delta h), \tag{23}$$

leading to the time delay

$$t_0 = \frac{\Delta h}{v_{ph}} \tag{24}$$

This means that the value of the field at the boundary can be predicted by buffering the fields one space step in front of the boundary for a time t_0. This time delay must be an integer multiple of Δt to fit all discrete conditions of the system, which demands the synchronism of pulses referring to the discretized space and time. Introducing an integer buffer depth

$$n_0 = \left[\frac{t}{\Delta t}\right] = \left[\frac{\Delta h}{v_{ph}}\right] = \left[2\sqrt{\varepsilon_{r,\mathit{eff}}}\right] \qquad (25)$$

leads to

$$_{n}E_{\tan}(I,j,k) = {}_{n-n_0}E_{\tan}(I-1,j,k) \qquad (26)$$

using the expressions for a discretized function as proposed by Yee [4]. This method can be applied to the 3D SCN TLM mesh provided that the plane of nodes in front of the boundary is guaranteed to support only TEM field propagation. The described algorithm works very well for a free space simulation, but instability occurs when inhomogeneous material fills the computational domain [22]. This instability problem can be solved by a modification of the algorithm, meaning that the field values are no longer predicted by buffering. Now the linear interpolation

$$_{n}E_{\tan}(I,j,k) = {}_{n-1}E_{\tan}(I,j,k) + \frac{1}{n_0}\left({}_{n-1}E_{\tan}(I-1,j,k) - {}_{n-1}E_{\tan}(I,j,k)\right) \qquad (8)$$

between two actual field values at the boundary $x_b = I \cdot \Delta h$ and one step Δh in front of it is used to predict the next field values at the boundary. Thus, it is no longer necessary to restrict the normalised value n_0, referring to the delay time t_0, to integer numbers. This algorithm usually named *1st order Mur* boundary condition [21] is stable and yield acceptable reflections.

It has been shown that on demand of the synchronism of pulses (referring to the discrete system of space and time) and in addition to the numerical dispersion of the used field theoretical method the absorbing boundary exhibits small reflections $\underline{r}$, means they do not work in an ideal manner.

Assume now a TEM- or quasi-TEM-waveguide to be considered by a transient type analysis. Then the voltage $\hat{\underline{u}}_{trans}$ and current $\hat{\underline{i}}_{trans}$ transmitted to the second port positioned at the distance ℓ consist of additional reflection terms, means

$$\hat{\underline{u}}_{trans} = (1+\underline{r}) \cdot \hat{\underline{u}}_{inc} \cdot e^{-j\beta\,\ell} \qquad \text{and} \qquad \hat{\underline{i}}_{trans} = (1+\underline{r}) \cdot \hat{\underline{i}}_{inc} \cdot e^{j\beta\,\ell} \qquad (28)$$

since the absorbing boundaries are assumed to work in an ideal manner.

A rigorous analysis yields that the small reflections from the absorbing boundaries result in an error of the scattering parameter, especially their phases [23]. This leads to a deviation in the apparent dielectric constant, because it depends on the square of the phase shift, with respect to the system variable (either currents or voltages) from which it has been calculated.
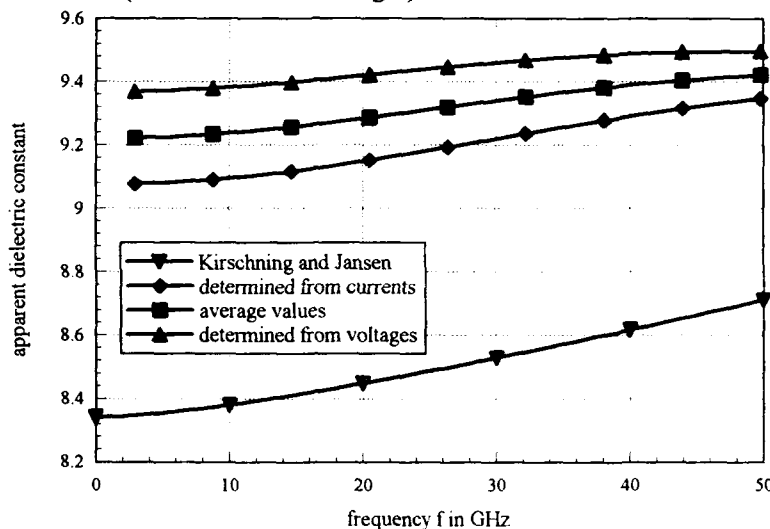

Figure 14: *The effective permittivity of a microstrip waveguide ($Z_0 = 50\Omega$, $h=100\mu m$, $w=75\mu m$, $t=0\mu m$) on a GaAs substrate is shown versus frequency: the offset to the values obtained from approximation formulas (Kirschning et. al. [24]) is due to coarseness error*

In Fig. 14 this deviation can clearly be seen. The apparent dielectric constant of a microstrip waveguide ($Z_0=50\Omega$) is plotted versus frequency f. The results obtained from the voltage phase-shift as well as from the current phase-shift and the average values of the phase-shifts are compared with well known approximation formulas (Kirschning et. al. [24]). In principle the curves have the same course while their slope is identical, whereby the offset to the results obtained from the approximation formulas is due to the coarseness error (the strip has been discretized by 6 elements). Further research has to be carried out to obtain better and stable ABCs for the TLM method or to adapt known FDTD ABCs to TLM.

5.1 Refined boundary positioning

The accurate modeling of waveguide components, discontinuities and junctions requires a precision in the positioning of boundaries that is identical to, or even better than the manufacturing tolerances. In traditional TLM models of electromagnetic structures, boundaries can only be placed either across the nodes [25] or halfway between nodes [12]. Unless all dimensions of the structure are integer multiples of the mesh parameter, the mesh parameter has to be very small indeed, leading to unacceptable computational requirements [26].

A method for changing the position of the boundaries in the two-dimensional TLM method through modification of the impulse scattering matrix of boundary nodes has already been described by Johns [27] in 1973. At that time, he thought that the advantage of this procedure over stepped contour modeling was too small to warrant the additional complexity of the algorithm.

The method proposed in [26] to overcome these problems leaves the scattering matrix intact, but replaces the simple boundary reflection coefficient r by a recursive reflection algorithm obtained from a rigorous transmission line treatment of the length extending half the length of the mesh parameter. Depending on the type of wall the transmission line extension is replaced by an inductance or a capacitance for an electric or a magnetic wall, respectively, as long as the wavelength of the highest propagating frequency is much larger than the mesh parameter. Solving the difference equations (discretized differential equations) for the total voltages and currents, and applying Eq. (9) with respect to the node the equation

$$_nU^i = \frac{1 - _n\kappa - (_n\kappa - _{n-1}\kappa)}{1 + _n\kappa + (_n\kappa - _{n-1}\kappa)} \cdot r + \frac{_n\kappa}{1 + _n\kappa + (_n\kappa - _{n-1}\kappa)} \cdot \left(_{n-1}U^r \cdot r + _{n-1}U^i\right), \qquad (29)$$

with $r = -1$ for an electric wall, $r = +1$ for a magnetic wall and the extension parameter $_n\kappa$ at time $_nt$ can be obtained. The positioning parameter $_n\kappa$ is equal to $2 \, \ell \, / \, \Delta h$ in the three-dimensional and $\sqrt{2} \, \ell \, / \, \Delta h$ in the two-dimensional TLM case.
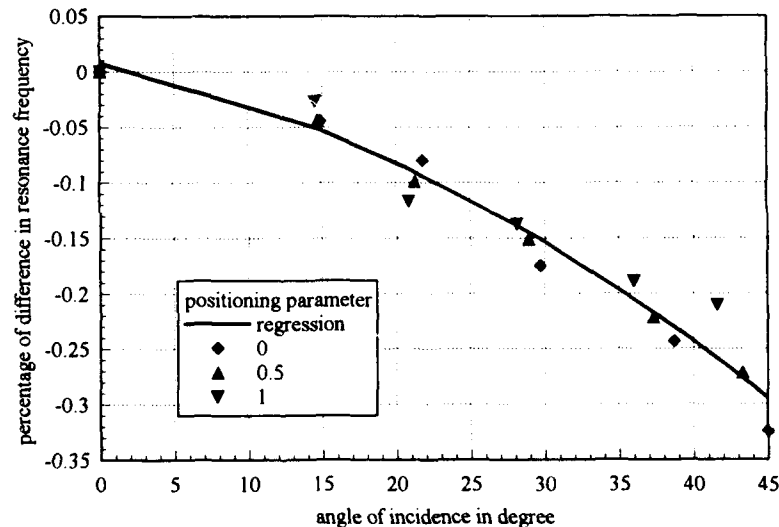

Figure 15: *The accuracy of the extended boundaries demonstrated for a resonant cavity: the percentage of the difference in the resonance frequency for a certain mode is plotted versus the angle of incidence of a plane wave corresponding to this mode.*

The accuracy of the above mentioned algorithm has been validated by performing extensive simulations of structures most sensitive to small variations in the dimensions, namely hollow

waveguide resonators. One of the walls was made moveable by application of Eq. (29) in combination with the SCN TLM scheme [10]. Fig. 15 demonstrates the results for the difference in the resonance frequency of the resonator obtained by the proposed algorithm in comparison with the theoretically predicted values. The difference in percent in plotted versus the angle of incidence for a plane wave corresponding to the evaluated mode in the resonator. Several computations are carried out for miscellaneous values of the positioning parameter, while the predictable error can be found from the regression curve. The largest error margin is obtained at an angle of incidence of about 45°.

Since the positioning parameter can be changed after each computational step by an arbitrary small amount, it becomes feasible to model effectively boundaries that move at arbitrary speed during the simulation. For instance, if the positioning parameter is programmed to increase linearly in time, the boundary appears to be moving at a constant speed away from its initial position. This allows us to model directly the Doppler shift in the time domain. Fig. 16 shows the effect of wall movement on the shape and delay of a Gaussian pulse as modeled with the two-dimensional TLM model. Two identical pulses of amplitude +1V have been reflected by electric walls (reflection coefficient r = -1) and are propagating towards the left at velocity c. One of the reflecting walls was stationary, and the other started to move away from the source at a constant speed v=0.035c at the moment of the incidence. The time delay, the pulse widening and the amplitude reduction due to the wall movement are clearly noticeable. The TLM results correspond exactly to the theoretical predictions and other numerical simulations [28].
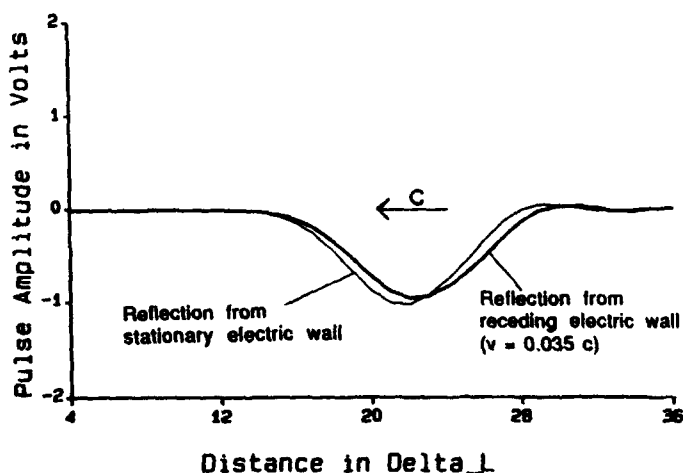
TLM MODELING OF DOPPLER SHIFT

Figure 16: *The influence of wall movement on the shape and delay of a Gaussian pulse as modelled with the two-dimensional TLM method. Two identical pulses of amplitude +1V have been reflected by electric walls and are propagating towards the left at velocity c.*

Due to the fact that this model enables only to treat extending boundaries more research activities must be spend even on the task of moving a boundary towards a node.

5.2 Reduction of Coarseness error

Like all other numerical techniques the TLM method is subject to various sources of error and must be applied with caution in order to yield reliable and accurate results [29]. It can be observed in all TLM computations that results are invariably shifted towards lower frequencies, particular when the structure under study contains field singularities. This phenomenon has already been described by Shih and Hoefer in the analysis of finlines with the two-dimensional TLM method [30]. The effect was called the coarseness error and attributed to the lack of resolution of highly nonuniform fields by the chosen discretization [29,30,31].

This error is an important source of inaccuracies in the analysis of planar structures containing regions like corners or edges. To reduce this error it is possible to choose a very fine mesh size, but this would lead to a demand of large memory and computation time, so that its accomplishment is at least impractical or even impossible. Therefore, other methods are necessarily introduced to overcome this problem. Hence, this led to the directional- or nondirectional compensation method after Mueller et. al. [32] as a coarseness error reduction measure.
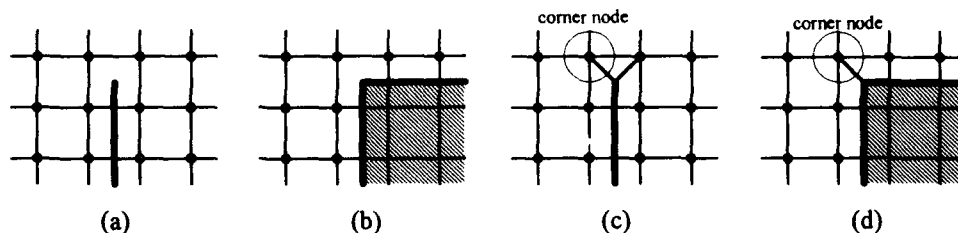

Figure 17: *2D TLM network containing (a) a 360° corner and (b) a 270° corner as well as the same structures containing the additional branches in (c) and (d), respectively [32].*

These methods base on the fact that at sharp corners of the cross section there are nodes, which are closer than Δh to the boundaries but have no branch directly connected to the boundary node as shown in Fig. 17a) and Fig. 17b). For directional compensation of the coarseness error a fifth branch, leading towards the boundary, is introduced as depicted in Fig. 17c) and Fig. 17d).

In the above referenced procedure, the admittances of the branches extending towards the corners or edges are modified. Not only the coordinates but also the relative position of the node with respect to the corner must be specified. Opposing the directional compensation, the method of nondirectional compensation adds an inductance to the center of the corner node

without respect to its relative position. Therefore, a stub of half mesh size length is added to the node without modifying the admittances of the other node branches. In this procedure only the coordinates of the node must be specified and the additional stub admittance has to be determined.

In both methods the corner node parameters, which describe the properties of the additional corner branches have been determined by optimising them for best compensation of the coarseness error in the resonance frequencies, hence, the problem of determining the parameters p (directional compensation) and y_5 (nondirectional compensation) still remains. Accurate resonance frequencies are obtained by a series of TLM simulations using increasingly fine meshes and extrapolating these results for $\Delta h \rightarrow 0$ (continuous function) as described in [30,31]. Then the same structures were computed with coarse meshes comprising corner nodes, and their parameters were optimised such that they yielded the accurate reference results.
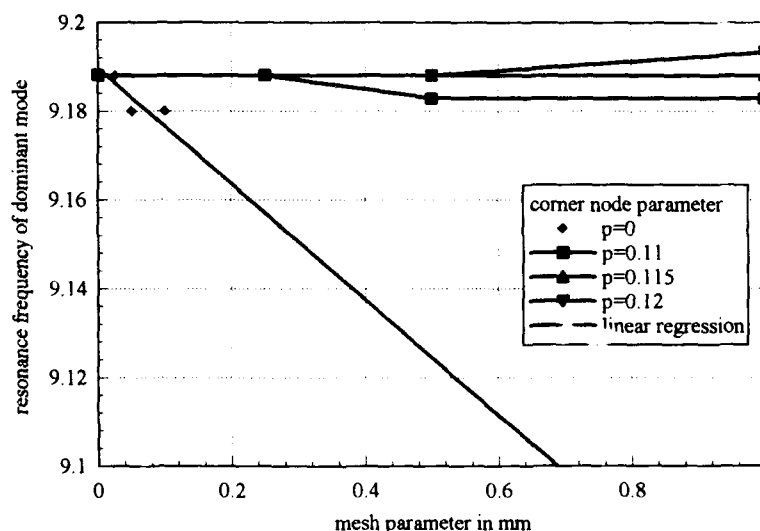

Figure 18: *Effect of a compensation mechanism on the example of the dominant mode resonance frequency for a quarter-wave resonator containing a 270° field singularity, while using the directional compensation [32].*

Fig. 18 gives examplarily the effect of the compensation mechanism on the dominant mode resonance frequency for a quarterwave resonator containing a 270° field singularity. For the simulations the directional compensation was used in order to reduce the coarseness error. The results were compared with accurate resonance frequencies calculated as proposed by Shih and Hoefer in [30,31]. It was found out that the corner node parameters were insensitive to mesh size, frequency and direction of propagation. The proposed techniques yield in all cases

excellent results with considerably less computational effort than uncompensated TLM schemes applying the same discretization [32].

5.3 Principles of nonlinear modelling using fields

Structures containing nonlinear elements can be modelled using different ways [33,34,35,36]. In order to take the real field distribution of the structure into account we can distinguish two different types of modelling:

- The compression approach after Kunisch et. al. [37,38], or,

- the embedding of the nonlinear device into the transient type field analysis [39].

Even the modelling of the device can be done differently. On the one hand, we can model the nonlinear device using its equivalent circuit directly represented in form of a difference equation (discretized differential equation) with time dependent coefficients [39,40] and on the other hand, we have the possibility to change the material properties of the nonlinear area in time with respect to the apparent capacitance and resistance [39,40,41]. The nonlinear differential equation can be integrated numerically either using the same time step as in the TLM field simulation or a piecewise linear integration algorithm with variable time step [41]. Both methods give good results as long as the voltage dependent elements of the nonlinear equivalent circuit vary only slowly from one time step to the next. This requirement guarantees that the values for the nonlinear elements can be calculated from the port voltage known from the previous time step [40], and it is usually fulfilled in order to have low numerical dispersion during the simulation.

6. Epilogue

Almost 300 years after Huygens' proposed his famous principle of wave propagation, this viewpoint has been rediscovered in order to carry out electromagnetic field computations. This principle was applied by Johns as a numerical fieldtheoretical method, known as the *Transmission Line Matrix* (TLM) method, and a variety of structures were calculated utilizing it. State of the art is to solve eigenvalue problems as well as initial boundary value problems by implementing the TLM method onto digital computers. Due to the invention of new field representation concepts as the SCN TLM method, the efficiency of TLM modelling has steadily approached that of the FDTD method. Even the development of faster serial as well as parallel computers pushed forward the research activities in the area of numerical TLM simulations for microwave and millimeter wave structures. Additionally, during the last decade a lot of effort was spent on the reduction of the main sources of error exhibited by the TLM method.

7. References:

[1] Hoefer, W.J.R: "Huygens and the computer - A powerfull alliance in numerical electromagnetics", *IEE Proceedings.*, Vol. 79, pp. 1459-1471, October 1991.

[2] Huygens, C.: "Traité de la Lumière", Leiden, 1690, Pierre vander Aa, reprinted in "Oevres Complètes de Cristiaan Huygens", Société Hollandaise des Sciences, vol. 19, Amsterdam, The Netherlands, Swets and Zeitlinger, 1967.

[3] Simonyi, K.: "*Kulturgeschichte der Physik*", (in German), Verlag Harri Deutsch, Thun, Frankfurt/Main 1990.

[4] Yee, K.S.: "Numerical Solution of initial boundary value problems involving Maxwell's equations in isotropic media", *IEEE Trans. Antennas Propagat.*, Vol. AP-14, pp. 302-307, May 1966.

[5] Johns, P.B.; Beurle, R.L.: "Numerical solution of 2-dimensional scattering problems using a transmission line matrix", *Proc. Inst. Elec. Eng.*, Vol. 118, pp. 1203-1208, Sep. 1971.

[6] Johns, P.B.: "The solution of inhomogeneous waveguide problems using a transmission line matrix", *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, pp. 209-215, March 1974.

[7] Akhtarzard, S.; Johns, P.B.: "Generalised elements for t.l.m. method of numerical analysis", *Proc. IEE*, Vol. 122, pp. 1349-1352, Dec. 1975.

[8] Collin, R.E.: "*Field theory of guided waves*", IEEE Press, New York 1991.

[9] Akhtarzard, S.; Johns, P.B.: "Solution of 6-component electromagnetic fields in three space dimensions and time by the TLM method", *Electron. Lett.*, Vol. 10, pp. 535-537, Dec. 12th 1974.

[10] Johns, P.B.: "A symmetrical condensed node for the TLM method", *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-35, pp. 370-377, April 1987.

[11] Uher, J.; Liang, S.; Hoefer, W.J.R.: "S-parameters of microwave components computed with the 3D condensed symmetrical node", *1990 IEEE MTT-S International Symposium Digest*, pp. 653-656, Dallas, USA, June 1990.

[12] Hoefer, W.J.R.; So, P.P.M.: "The electromagnetic wave simulator", J Wileys & Sons, Chichester 1991.

[13] Mueller, U.; Rittweger, M.; Beyer, A.: "Coplanar short considered by the TLM-method with symmetrical condensed nodes", *Proceedings of the 21st European Microwave Conference 1991*, pp. 999-1003, Stuttgart, GERMANY, Sept. 1991.

[14] Saguet, P.; Pic, E.: "An improvement of the TLM method", *Electron. Lett.*, Vol. 16, pp 247-248, March 27th 1980.

[15] Wills, J.D.: "Spectral estimation for the transmission line matrix method", *IEEE Trans. Theory Tech.*, Vol. MTT-38, pp. 448-451, Apr. 1990.

[16] Mueller, U.; Beyer, A.; Rittweger, M.: "Simulations with the 3D TLM SCN using FD-TD absorbing boundary conditions", *1992 MTT-S International Symposium Digest*, Vol. 1, pp. 377-380, Albuquerque, USA, June 1992.

[17] Octopus 1.7: Network- and noise analysis program and manual, ArguMens Mikrowellentechnik GmbH, Duisburg, Germany.

[18] **Moore, T.; Blaschak, J.; Taflove, A.; Kriegsmann, G.:** "Theory and application of radiation boundary operators", *IEEE Trans. Antennas Propagat.*, Vol. AP-36, pp 1797-1812, Dec. 1988.

[19] **Simons, N.R.S.; Bridges, E.:** "Application of absorbing boundary conditions to TLM simulations", *1990 IEEE AP-S International Symposium Digest*, Vol. 1, pp. 2-5, Dallas, USA, May 1990.

[20] **Saguet, P.:** "TLM method for the three dimensional analysis of microwave and MM-wave structures", *Workshop Proceedings IEEE MTT/AP German Chapter*, pp. 99-104, Stuttgart, Sep. 1991.

[21] **Mur, G.:** "Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equation", *IEEE Trans. Electromagnetic Comp.*, Vol. EMC-23, pp. 377-382, Nov. 1981.

[22] **Rittweger, M.; Wolff, I.:** "Analysis of complex passive (M)MIC-components using the finite difference time domain approach", *1990 IEEE MTT-S International Symposium Digest*, Vol. 3, pp. 1147-1150, Dallas, USA, June 1990.

[23] **Mueller, U.; Rittweger, M.; Beyer, A.:** "Influence of absorbing boundaries in time domain iterative methods", *Proceedings of the 22nd European Microwave Conference 1992* ,pp. 1001-1005, Espoo, FINLAND, Aug. 1992.

[24] **Kirschning, M.; Jansen, R.:** "Accurate model for effective dielectric constant of microstrip with validity up to the millimetre-wave frequencies", *Electron. Lett.*, Vol. 18, pp. 272-273, March 1982.

[25] **Chen, Z.; Ney, M.; Hoefer, W.J.R.:** "A new boundary description in two-dimensional TLM models of microwave circuits", *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-39, No. 3, pp. 377-382, March 1991.

[26] **Mueller, U.; Beyer, A.; Hoefer, W.J.R.:** " Moving boundaries in 2-D and 3D simulations realized by recursive formulas", *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-40, No. 12, pp. 2267-2271, Dec. 1992.

[27] **Johns, P.B.:** "Transient analysis of waveguides with curved boundaries", *Electron. Lett.*, Vol. 9, No. 21, Oct. 18, 1973.

[28] **Harfoush, F.; Taflove, A.; Kriegsmann, G.A.:** "A numerical technique for analyzing electromagnetic wave scattering from moving surfaces in one- and two dimensions", *IEEE Trans. Antennas Propagat.*, Vol. AP-37, pp. 55-63, Jan. 1989.

[29] **Hoefer, W.J.R.:** "The transmission line matrix (TLM) method", in Itoh, T.: *"Numerical techniques for microwave and millimetre-wave passive structures"*, J. Wiley & Sons, New York 1989.

[30] **Shih, Y.-C.; Hoefer, W.J.R.:** "Dominant and second-order mode cuttoff frequencies with a twodimensional TLM-program", *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, pp. 1443-1448, Dec. 1980.

[31] **Shih, Y.-C.:** "The analysis of fin lines using transmission line matrix and transverse resonance method", *M. A. SC. Thesis*, University of Ottawa, Canada, 1980.

[32] **Mueller, U.; So, P.P.M; Hoefer, W.J.R.:** "The compensation of coarseness error in 2D TLM modelling of microwave structures", *IEEE 1992 MTT-S Symposium Digest*, Vol. 1, pp. 373-375, June 1992.

[33] **Nagel, L.W.; Pederson D. O.:** "SPICE (Simulation Program with Integrated Circuit Emphasis)", University of California, Electronics Research Laboratory, Memorandum ERL-M382, April 1973.

[34] **Shelboe, S.:** "Computations of the periodic steady state response of nonlinear networks by extrapolation methods", *IEEE Trans. Circuit Syst.*, Vol. CAS-27, pp. 161-175, March 1980.

[35] **Apsille, T.J.; Trick T.N.:** "Steady-state analysis of nonlinear circuits with periodic inputs", *Proc. IEEE*, Vol. 60, pp. 108-114, Jan. 1972.

[36] **Kundert, K.; Sangiovanni-Vicentelli, A.:** "Simulation of nonlinear circuits in the frequency domain", *IEEE Trans. Computer Aided Design*, Vol. CAD-5, pp. 521-535, May 1986.

[37] **Wolff, I.; Kunisch, J.:** "The compression approach: A new technique for the analysis of distributed circuits containing nonlinear elements", *1992 IEEE MTT-S International Microwave Symposium Workshop Digest*, Workshop on "CAD of nonlinear microwave circuits using field theoretical methods", June 1992.

[38] **Kunisch, J.; Bahr, A.; Rittweger, M.; Heinen, S.; Wolff, I.:** "Analysis of nonlinear forced and autonomous microwave circuits using the compression approach: Theory and application to a nonlinear transmission line", *1992 INMMC Workshop Digest*, pp. 289-300, Duisburg, GERMANY, Oct. 1992.

[39] **Russer, P.; So. P.P.M; Hoefer, W.J.R.:** "Modelling of nonlinear active regions in TLM", *IEEE Microwave Guided Wave Lett.*, Vol. MGWL-1, pp. 10-13, January 1991.

[40] **Hoefer, W.J.R.; Nielsen J.:** "Time domain nonlinear circuit analysis using the TLM method", *1992 IEEE MTT-S International Microwave Symposium Workshop Digest*, Workshop on "CAD of nonlinear microwave circuits using field theoretical methods", June 1992.

[41] **Voelker, R.H.; Lomax, R.J.:** "A finite-difference transmission line matrix method incorporating a nonlinear device model", *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-38, pp. 302-312, March 1990.

MICROWAVE REMOTE SENSING
OF ROAD SURFACE DURING WINTER TIME

Gottfried MAGERL and Werner PRITZL

Abstract - We report a bistatic microwave radar operating at 2.45GHz for the contactless evaluation of weather induced road conditions. The radar determines by an ellipsometric method whether the road surface is dry, wet, snow or slush covered. In the case of a wet road the evaluation of thickness, of salinity, and from that of the freezing point of the water layer is possible with $\pm 2^{\circ}C$ accuracy for layer thicknesses of at least 0.07mm. We describe the hardware, in particular the microstrip antenna design and the layout of the receiver. We present typical results of measurements and we discuss principal constraints of the method when either the water layer or the road material freezes. The behavior of the radar for inhomogeneous layers on the road is evaluated, too.

I. INTRODUCTION

The safety of car driving critically depends on weather induced road conditions. This is particularly true during winter time. Therefore, several attempts have been made to remotely sense the road surface by the use of microwave radar devices: W. Hetzner investigated the possibility of a dry- or wet-road decision by a 50-GHz reflectometer to be installed on the car [1]. S. Hertl and coworkers developed a bistatic radar at 10 GHz for measuring the thickness of a water-layer covering the road surface [2]. H. Scharsching tested the suitability of this device for application on highways [3]. He pointed out that the additional knowledge of the salinity of the water and, therefore, of its freezing point would be extremely helpful for road maintainance during winter time. At this point we started the development of a stationary road condition radar suitable for determining whether a road would be dry, wet, or covered by snow or slush, and additionally capable of measuring thickness and salinity of water layers.

In the following Section II we discuss the principle of the measurement, in Section III we give a brief description of the microwave hardware, in Section IV we present results of measurements, and in Section V we discuss constraints of the method.

II. PRINCIPLE OF MEASUREMENT

Figure 1 shows schematically a plane electromagnetic wave that travels through a homogeneous medium (1) with (real) dielectric constant ε_1- in our case air - and impinges at an angle of incidence ϑ_1 onto a layered structure. This structure consists of a layer of medium (2) with (complex) dielectric

---

The authors are with the Institut für Nachrichtentechnik und Hochfrequenztechnik, Technische Universität Wien, Gußhausstraße 25 / 389, A-1040 Wien, Austria.

constant ε_2 and thickness d covering medium (3) with (complex) dielectric constant ε_3 and infinite thickness - the road. In general, we will have both a p- (parallel or vertically) and an s- (perpendicularly or horizontally) polarized component. It can be easily shown [4] that the complex reflection coefficient of the electric field strength is given by

$$R_{p,s} = \frac{R_{12} + R_{23}\,exp(-2\,j\beta)}{1 + R_{12}\,R_{23}\,exp(-2\,j\beta)} \qquad (1)$$

with

$$\beta = 2\pi\frac{d}{\lambda}\sqrt{\varepsilon_2}\,cos\vartheta_2, \qquad (2)$$

where the reflection coefficient R_{12} between media 1 (air) and 2 (road covering layer, e.g. water) has to be inserted from Fresnel's formulae [5] for p- and s-polarized radiation depending on whether R_p or R_s has to be determined. It should be pointed out that the unknown reflection coefficient R_{23} (layer-road) can be substituted by R_{13} (air-road) and R_{12} (air-layer) according to

$$R_{23} = \frac{R_{13} - R_{12}}{1 - R_{12}R_{13}}, \qquad (3)$$

where R_{13} may be determined by a calibration procedure when installing the road condition radar on a dry day. As will be shown in Section V, it is extremely important to recalibrate R_{13} whenever possible, in particular at temperatures below freezing, as the dielectric properties of the road surface may change dramatically with changing temperature. If the complex reflection coefficients R_p and R_s can be measured, it is possible to solve eqs. (1) and (2) for the unknown quantities d and ε_2. Thus, thickness and kind of layer covering the road can be evaluated, as water, snow, and slush differ sufficiently in their respective dielectric constants in order to be distinguished from each other [6-8].
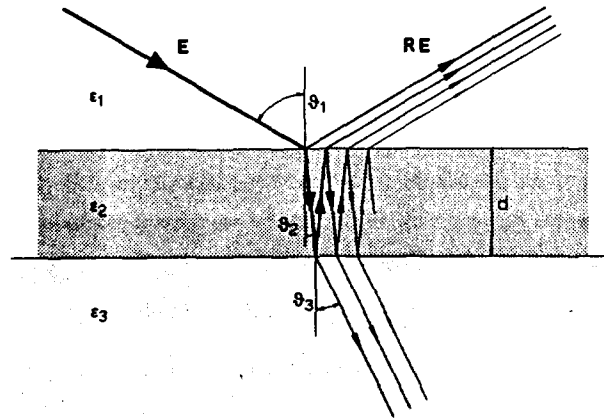

Fig. 1: Schematic of a plane electromagnetic wave with electric field strength E being reflected at a layer of thickness d and (complex) dielectric constant ε_2 covering the road surface with (complex) dielectric constant ε_3; the resulting complex reflection coefficient is denoted by R.

The direct measurement of complex reflection coefficients, in particular the detection of their absolute phase angles is not easily and accurately feasible in the road environment where we have to deal with distances between transmitter and receiver of roughly 100 wavelengths. However, a closer examination reveals that it is sufficient to evaluate just the magnitudes $|R_p|$ and $|R_s|$, where we need for most of our calculations the **ratio** $|R_p/R_s|$, and the phase **difference** $\arg(R_p/R_s)$. Besides of simplicity, the relative measurement has the advantage of common-mode rejection: Fluctuations of transmitted microwave power and variations in the phase constant of the transmission path through the air due to fog, rain, or snow are cancelled out.

III. MICROWAVE HARDWARE DESCRIPTION

The hardware of the radar consists of a transmitter operating at the ISM-Frequency of 2.45GHz and of a corresponding receiver, both mounted at a height of about 3m above ground on opposite sides of the road to be monitored. The geometry is chosen to provide an angle of incidence $\vartheta_1 \sim 60^{\circ}$ which is close to the Brewster angle of the air-road interface and gives about optimum contrast in the measured signals when the road gets wet. The transmitter is a simple dielectrically frequency-stabilized transistor oscillator which feeds an electronically controllable output power in the range of 5mW to 50mW into the transmitter antenna. The antenna is a planar microstrip array of four rows of four elements each which are fed in parallel via appropriate microstrip feeding lines and T-junctions. The 3-dB width of its main lobe is about 25° such illuminating an area on the road surface of roughly $10m^2$. The antenna radiation pattern has a zero at about 30° such suppressing direct-line-of-sight connection and corresponding interference between transmitting and receiving antennas. Within the planar group, the feeding currents are weighted 0.63 : 1.0 : 1.6 to achieve a sidelobe suppression of better than 15dB. The transmitter antenna is rotated by 45° about its surface normal to provide equal amounts of p- and s-polarized electromagnetic radiation.

The receiver consists of two microstrip antenna arrays of the same construction as the transmitter antenna, mounted side by side and arranged such that one array picks up just the p-polarized component of the incoming wave and the other one just the s-polarized component, respectively. Since we measure the phase difference between these two components, any angular misalignment of the two receiving antennas - in particular a rotation about the vertical axis of the mounting pole - causes a corresponding error in the measurement. As an example, a $\pm 1mm$ tilt sideways (at the outmost edges of the arrays!) due to such a rotation about the central vertical axis of symmetry corresponds to an error of 3° in the phase difference. This extreme sensitivity of alignment turned out to be a severe drawback during installation and calibration of the receiver, and later on during operation under heavy windloads. To circumvent this problem, we developed a receiving antenna with both sub-groups folded into each other with almost coinciding phase centers. According to calculations, the phase error due to angular misalignment should be a tenth of the previous value. However, this new generation of receiving antennas is not yet installed, therefore, we don´t have practical experiences so far.
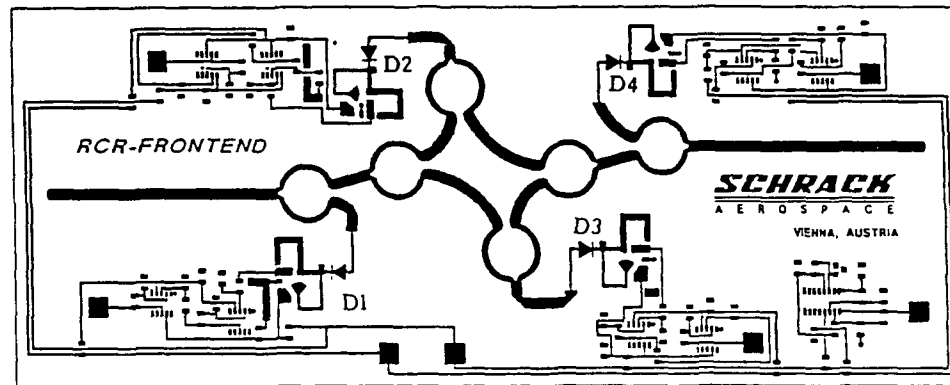
Fig. 2: Print layout of the front end of the road condition radar receiver.

The two antenna signals are fed into a microstrip front end. Its print layout is shown in Fig. 2. Essentially, the front end is a sixport type homodyne receiver where the two incoming signals proportional to R_p and to R_S are beat against each other. In a first step, each input signal is split by a Wilkinson-type 3-dB power divider feeding one part of each signal into detector diode $D1$ and $D4$, respectively. These diodes provide dc-signals V_1 and V_4 proportional to $|R_p|^2$ and $|R_s|^2$

$$V_1 = 0.5 \times C \times |R_p|^2 \tag{4}$$

$$V_4 = 0.5 \times C \times |R_s|^2. \tag{5}$$

In a second step, the remaining parts of the input signals are split again and recombined with each other by Wilkinson combiners via delay lines providing a mutual phase difference of 90°. Thus, the dc signals V_2 and V_3 provided by diodes $D2$ and $D3$ are proportional to a sum of V_1 and V_4 and of the sine and the cosine of $\Phi = \arg(R_p) - \arg(R_s)$

$$V_2 = 0.125 \times C \times \left(|R_p|^2 + |R_s|^2 + 2|R_p| \times |R_s| \times \cos\Phi\right) \tag{6}$$

$$V_3 = 0.125 \times C \times \left(|R_p|^2 + |R_s|^2 - 2|R_p| \times |R_s| \times \sin\Phi\right). \tag{7}$$

The constant C denotes the sens ivity of the four (identical) detector diodes, and the factors 0.5 and 0.125 take the splitting of the signals into account. The diodes are kept at constant (elevated) temperature, and both the transmitter and the receiver are mounted in protective housings with microwave transparent plastic covers in front of the antennas. To avoid difficulties with small dc-signals embedded in 1/f-noise, we apply 100% AM to the transmitted 2.45-GHz signal at a modulation frequency of 1kHz. This makes frequency selective detection of the four diode signals possible which is followed by A/D-conversion. Finally, a microcomputer evaluates $|R_p|$, $|R_s|$, and $\arg(R_p/R_s)$ from the digitized signals by the use of eqs. (4) to (7), and from that the unknown quantities d, ϵ_2 and the salinity are determined by numerically solving eqs. (1) to (3).

IV. MEASUREMENT RESULTS

The absolute value of the amplitude ratio $|R_p/R_s|$ and the corresponding phase difference may be visualized in the complex plane as constant-angle-of-incidence contours [4]. Figure 3 shows the theoretical dependence of the real part $Re(R_p/R_s)$ and of the imaginary part $Im(R_p/R_s)$ on the thickness d of the water layer with its freezing temperature serving as a parameter. All contours start at a point corresponding to the dry road surface, extending to the left for increasing water film thickness d. The contour for pure water exhibits the smallest (negative!) imaginary part, whereas the curves for salt (NaCl) water approach the real axis with increasing salinity corresponding to decreasing freezing temperatures. Therefore, any water layer on the road with given thickness and concentration of salt corresponds to a single point in the diagram.
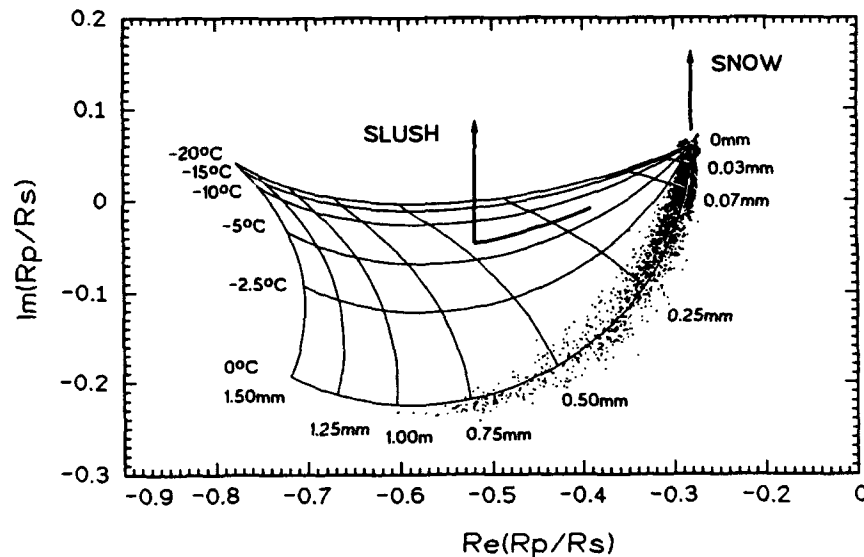

Fig. 3: Constant-angle-of-incidence contours ($\vartheta_1 = 60^\circ$) of salt water layers on a concrete road surface; parameters are thickness (d = 0, 0.03, 0.07, 0.25, 0.50, 0.75, 1.00, 1.25, and 1.50mm) and freezing temperature (0, -2.5, -5, -10, -15, and -20°C) derived from the salinity of the water film. Measurements, represented by dots, were taken from Oct.1 to Oct.21, 1993, on highway A10.

From Fig. 3 one can clearly see that the measurements represented by dots closely follow the theoretical "pure-water line" as can be expected in the month of October. However, we observe a considerable scatter in the measured values. This prohibits the evaluation of a (somewhat reliable) value of salinity at water layer thicknesses less than 0.07mm.

If snow covers a dry road, the snow layer acts like an antireflection coating due to the dielectric constant of (dry) snow lying just in between the permittivities of air and of the road material, respectively. In this case, the dry point of the road is shifted upwards as indicated by the arrow in Fig.3. An additional criterion for our road condition radar system to announce "snow on the road" is the absolute value of the reflection coefficient of the s-polarized component being smaller than that of the dry road $|R_s| < |R_{s,dry}|$ and the temperature of the road surface being less than $+1^oC$. If snow is falling on a wet road and slush develops on the surface, the measured values move in upward direction starting from the corresponding wet point and ending up outside the theoretical "wet area" as already shown in [6] and indicated in Fig.3, too.

A still unsolved problem is the unambiguous detection of an icy road. The dielectric constant of ice is too low ($\varepsilon_r = 3.14$) to produce a reflection sufficiently different from that of the dry road. This fact excludes the detection of ice building up on the road surface by sublimation of humidity out of the air. Even the detection of a freezing salt water layer is difficult as can be seen from Fig.4: The full line gives the temporal behavior of the freezing point of the humid and salty road surface as derived from the measurement according to Fig.3. Again, an appreciable scatter in the measured values is present. The broken line represents the road surface temperature as determined by a thermocouple of type PT100. From 6.00 a.m. until about 8.30 a.m. the road temperature is almost constant at -0.5^oC and the freezing temperature of the water on the road scatters around -1.5^oC. From 9.00 a.m. to 7.00 p.m. one sees the road temperature rising to $+4^oC$ and falling back again to 0^oC, whereas the freezing point
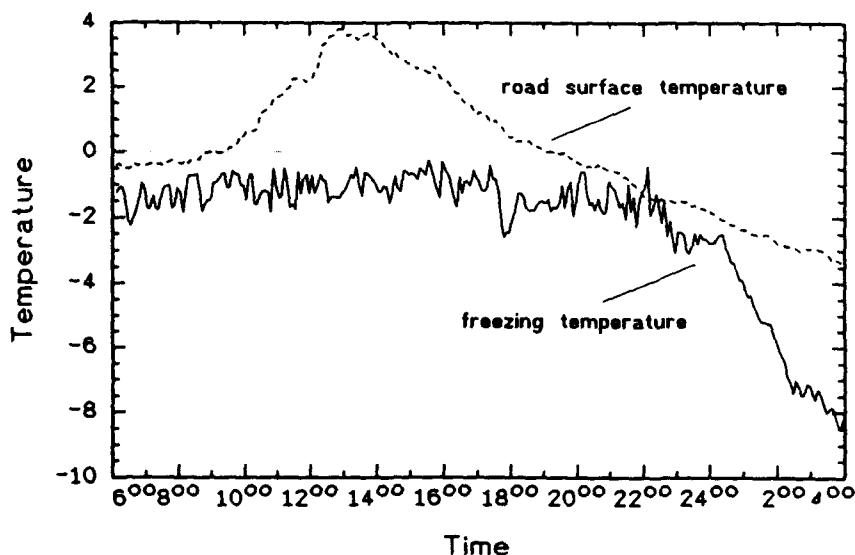

Fig. 4: Measured freezing point (full line) and road temperature (broken line).

remains unchanged with the exemption of a short drop to -2.5°C due to the application of salt at around 6.00 p.m. Around 10.00 p.m. road temperature and freezing temperature become identical, and since then the freezing temperature drops dramatically down to -8°C with no further application of salt reported in the road maintainance log. The reason for this phenomenon is freezing of the salt water on the road which results in a mixture of pure ice crystals (-undetected by the microwave) and of residual water with a higher salt concentration which is clearly detected by the radar. Therefore, the apparently safe condition "freezing point far below road temperature" in reality is a very dangerous situation. This can only be handled by a sophisticated software taking into account the temporal development of the measured values and performing plausibility checks.

V. CONSTRAINTS OF THE METHOD

In this Section we will discuss two principal constraints of the road condition radar measurements which are imposed by natural occurence. The first constraint arises when the road freezes: The dielectric constant ε_3 of the road is given by a weighted mean value of the permittivities of the pure road material - in most cases concrete or asphalt - and, due to its porosity, of its water content. When temperature drops below freezing, the latter relative dielectric constant drops accordingly from $\varepsilon_r = 81.5 - j20.7$ (water) to $\varepsilon_r = 3.14$ (ice) thus appreciably lowering the value for the road. As a consequence, the "dry point" of the road is shifted in a rather unpredictable way. We solved the problem by frequent recalibration of the dry road surface which means a determination of R_{13} (comp. eq.2) taking into account changing values of ε_3. Again, this is only possible with a software which is able to discern a dry road surface below freezing temperature by performing plausibility checks.

Another phenomenon which can be observed on highways is one lane being still wet when the other lane is already drying. Therefore, we investigated what happens if a radar "looks" about equally at both lanes. In a first step we modelled the surface of the road by squares of 5cm side length each and we calculated the contribution of each single surface element to the resulting signal. For this purpose we calculated the electric field strength of the incoming wave taking into account the radiation pattern of the transmitter antenna and the length of the path of propagation. The reflection was calculated according to eqs.1 - 3, additionally attributing an individual radiation pattern to each element, calculated using the Huygens' principle. Then the signal received from the given element was calculated considering the length of the air path to and the pattern of the receiving antenna. Finally, the contributions of all surface elements were added. As a welcome by-product this procedure gave us valuable insights with respect to the validity of the plane-wave approximation and with respect to what area of the road surface substantially contributes to the measurement results. In order to investigate our original problem of uneven distribution of water on the road surface, we assumed a uniform water layer of 0.25mm thickness from which we subtracted a "water step" of 0.18mm. As shown in the inset of Fig.5, we assumed this step to propagate across the width of the highway starting from the receiver side and ending up at the transmitter side resulting in an again uniform layer of a (reduced) thickness of 0.07mm.
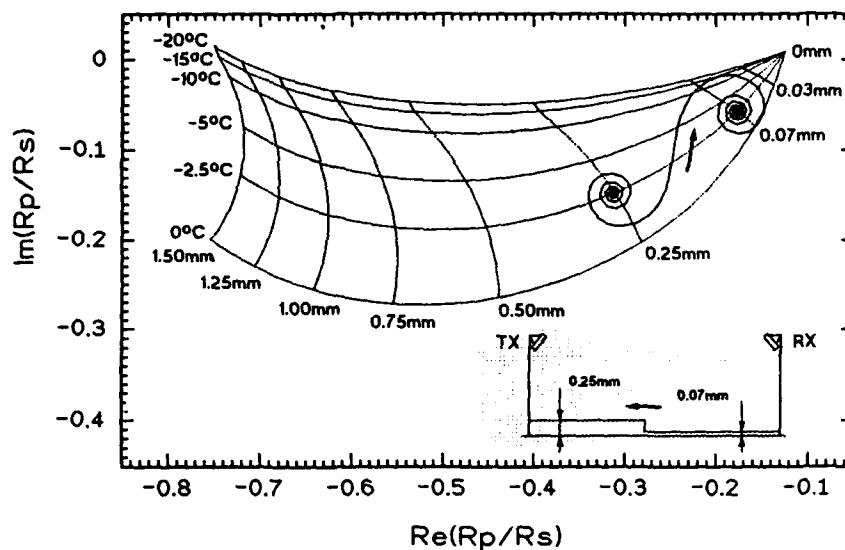
Fig. 5: Computer simulation of a "step" of salt water of 0.18mm height on top of a homogeneous layer of 0.07 mm thickness. This step is shifted across the width of the highway from the receiver to the transmitter side.

Figure 5 shows the result of our calculation: When the inhomogeneity in the thickness of the salt water layer is close to the receiver, we observe a result close to the road covered by the thick layer. With the step propagating towards the middle of the road, the resulting values spiral rather closely around the starting point. Once the step is in the middle of the road, the Cornu's spiral rather rapidly sweeps through the diagram and widely encircles the point of the thin layer which has a detrimental effect in particular on the "measured" salt concentration. However, in this case of an inhomogeneous layer on the road surface there is no software remedy to the problem.

CONCLUSIONS

We designed and constructed a bistatic microwave radar which allows to remotely sense the weather induced condition of road surfaces. The radar applies an ellipsometric measurement method and it detects whether a road is dry, wet, snow or slush covered. In the case of a wet road it additionally determines thickness and salinity of the water layer. We discussed some important features of the microwave hardware, in particular we presented the receiver and showed the evaluation of quantities related to the reflection of the electromagnetic wave at the road surface. We presented measurement results and discussed their accuracy. Finally, we showed problems of the unambiguous detection of an icy road, and we evaluated some principal constraints of the method connected to freezing road material and to inhomogeneous layers on the road.

REFERENCES

[1] W. Hetzner, Aktive und passive Straßenzustandserkennung im Millimeterwellenbereich, Frequenz, vol.38, pp.179-185, July / Aug. 1984.

[2] S. Hertl, G. Schaffar, and H. Störi, Contactless determination of the properties of water films on roads, J. Phys. E: Sci. Instrum., vol.21, pp.955-958, Oct. 1988.

[3] H. Scharsching, Berührungslose Feststellung der Fahrbahnnässe und Anwendung für Verkehrsbe-einflussungsanlagen, Doctor's Thesis, Technische Universität Wien, Oct. 1986.

[4] R.M.A. Azzam and N.M. Bashara, Ellipsometry and polarized light, Amsterdam: North Holland Publishing Company, 1977 Ch.4.3, pp.283 - 288.

[5] L. Solymar, Lectures on electromagnetic theory, Oxford: University Press, 1984, Ch.5.15, pp.166 - 170.

[6] G. Magerl, W. Pritzl, and P.W. Fröhling, Remote sensing of road condition, Proc. 1991 Internat. Geoscience and Remote Sensing Symp. (IGARSS'91), vol.IV, pp.2137-2140.

[7] W. Pritzl, G. Magerl, and P.W. Fröhling, A road-condition sensing microwave radar, Proc. 1991 Internat. Symp. on Recent Advances in Microwave Technology (ISRAMT'91), pp. 376-381.

[8] P.W. Fröhling, G.Magerl, and W. Pritzl, Detection of weather induced road condition, Proc. 3rd Int. Symp. Snow Removal and Ice Control Technology, 1992, vol. 2, preprint #34.

AUTHOR INDEX